On the Use of an Adaptive, Hybrid-Isentropic Vertical Coordinate in Global Atmospheric Modeling

RAINER BLECK

*NASA Goddard Institute for Space Studies, Columbia University, New York, New York, and
NOAA/Earth System Research Laboratory, Boulder, Colorado*

STAN BENJAMIN, JIN LEE, AND ALEXANDER E. MACDONALD

NOAA/Earth System Research Laboratory, Boulder, Colorado

ABSTRACT

This article is one in a series describing the functionality of the Flow-Following, Finite-Volume Icosahedral Model (FIM) developed at NOAA's Earth System Research Laboratory. Emphasis in this article is on the design of the vertical coordinate—the ''flow following'' aspect of FIM. The coordinate is terrain-following near the ground and isentropic in the free atmosphere. The spatial transition between the two coordinates is adaptive and is based on the arbitrary Lagrangian–Eulerian (ALE) paradigm. The impact of vertical resolution trade-offs between the present hybrid approach and traditional terrain-following coordinates is demonstrated in a three-part case study.

1. Introduction

The last few decades have seen significant advances in numerical weather prediction (NWP). The skill of today's NWP models owes much to improved closure schemes for physical processes that are too short lived or too small in scale to be resolved by a model's space–time mesh. Higher numerical accuracy, made possible primarily by faster computers but to some degree by new techniques for solving partial differential equations, also had a large impact on forecast skill.

Numerical accuracy is typically expressed in terms of the truncation or discretization error, defined as the extent to which individual terms in a differential equation are misrepresented in a numerical model because of grid resolution limits. There are at least three ways to reduce the truncation error: one can 1) refine the grid, 2) approximate spatial and temporal derivatives in a more precise way, or 3) optimize the placement of grid points in the space–time domain.

The NWP Flow-Following, Finite-Volume, Icosahedral Model (FIM) that was recently developed by the National

Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) takes advantage of recently developed ideas about gridpoint placement—the third alternative just mentioned. Specifically, FIM uses the icosahedron, a near-spherical body composed of 20 equilateral triangles, as a basis for horizontal grid layout, while in the vertical it uses a coordinate with a strong Lagrangian (hence ''flow following'') flavor. The focus of the present article is on the vertical coordinate in FIM.

2. History

Physical reasoning suggests that entropy or a variable related to it, such as buoyancy, is the most appropriate candidate for a Langrangian vertical coordinate in modeling stratified, quasi-adiabatic flow. The relationship between buoyancy and available potential energy assures that isentropic coordinate layers cannot steepen without a source of energy; this provides an important, although certainly not universal, safeguard against folding of coordinate surfaces that would otherwise bring the simulation to a quick halt.

The advantages of analyzing atmospheric motion in the free atmosphere in an isentropic reference frame have been thoroughly discussed in the literature starting with

Corresponding author address: Dr. Rainer Bleck, NOAA/ESRL, Global Systems Division, 325 Broadway, Boulder, CO 80303.
E-mail: rainer.bleck@noaa.gov

Rossby and Collaborators (1937), Rossby (1940), and Kleinschmidt (1950). Their arguments need not be repeated here. Likewise, the rationale for using isentropic coordinates in NWP models has been laid out repeatedly (e.g., Eliassen and Raustein 1968; Bleck 1974; Hsu and Arakawa 1990; Benjamin et al. 2004). The list of potential advantages of isentropic modeling compiled by those and other authors is long, and there is not much we can add to it at this time.

Isentropic NWP models came into being in the 1960s, but it is fair to say that "pure" isentropic coordinate models (those that use entropy as vertical coordinate throughout the model domain) have not withstood the test of time because of their inherent inability to provide vertical resolution in unstably stratified air columns. Complexities associated with coordinate–ground intersections also were, and continue to be, a deterrent to pure isentropic coordinate modeling.

Early experiments with isentropic models were conducted to simulate baroclinic instability with an eye on short-range weather prediction. Since modeling of diabatic forcing was not essential in that context, the intersection of coordinate surfaces with the ground—unavoidable in baroclinic flow—was regarded as the main numerical challenge. Eliassen and Raustein (1968), in their pioneering work on primitive equation isentropic modeling, chose to track coordinate–ground intersections by solving an advection equation for surface potential temperature θ_s:

$$\frac{\partial \theta_s}{\partial t} + \mathbf{v}_s \cdot \nabla \theta_s = 0. \tag{1}$$

(Here, $\mathbf{v}_s$ is the surface wind vector and ∇ is the two-dimensional gradient operator.) Although justifiable at the time, this strategy created a redundancy problem because isentropic coordinate models typically contain a second equation predicting the location of coordinate–ground intersections, namely, the continuity equation for isentropic layer thickness Δp:

$$\frac{\partial \Delta p}{\partial t} + \nabla \cdot (\mathbf{v} \Delta p) = 0. \tag{2}$$

To understand why this is so, one must keep in mind that the line marking the intersection of an isentrope with the ground also marks the edge of the region where the thickness Δp of the coordinate layer beneath it is zero. The evolution of the Δp field in the vicinity of the intersection line therefore provides information about where the line is moving.

Obtaining accurate solutions of (2) in the transition region between zero and nonzero Δp values is numerically challenging. Since the solution of (1) is subject to

numerical errors as well, the two equations do not always agree on where a coordinate surface intersects the ground at any given time. The resulting discrepancies act at best as a source of model noise; at worst, they lead to numerical instability.

The problem just described spawned attempts in the 1970s to improve the treatment of coordinate–ground intersections by inserting a set of terrain-following coordinate surfaces, commonly referred to as σ surfaces (Phillips 1957), between the ground and the isentropic domain (Deaven 1976; Uccellini et al. 1979; Friend et al. 1977; Bleck 1978b). In all of these schemes, with the exception of scheme D of Bleck (1978b), the two coordinate domains overlap (i.e., coordinate surfaces belonging to one domain intersect those belonging to the other). This requires interpolation. Scheme D, having no overlaps, leads to a particularly simple set of model equations, making it easy to formulate them in rigorously conservative form, but it has shortcomings of its own. The Uccellini et al. (1979) scheme and a variant of Bleck's scheme D are in use today in different versions of the University of Wisconsin global model (Zapotocny et al. 1994; Schaak et al. 2004).

The decade of the 1980s saw some progress in the related field of ocean circulation modeling with an entropy-related vertical coordinate. Specifically, Bleck and Boudra (1981) developed a coordinate system that is mainly isopycnic but allows coordinate layers to turn into constant-thickness layers near the sea surface to overcome the massless-layer problem associated with modeling baroclinic ocean states. This may have been the first time that arbitrary Lagrangian–Eulerian (ALE)-like coordinates (Hirt et al. 1974) were used in geophysical modeling.

Increasingly skillful schemes for avoiding coordinate–ground intersections were developed in the 1990s (Bleck and Benjamin 1993; Konor and Arakawa 1997; Johnson and Yuan 1998; Webster et al. 1999). These more recent schemes resemble scheme D of Bleck (1978b) in the sense that they avoid overlaps between a purely isentropic and a purely isobaric or terrain-following subdomain. They differ among each other primarily in how spatially abrupt they allow the transition between non-isentropic and isentropic coordinate representation to be. Various aspects of these schemes will be discussed later in greater detail.

A parallel effort to build purely isentropic or isopycnic models free of the redundancy implied by jointly solving (1) and (2) was brought to a conclusion (Bleck 1984; Bleck and Boudra 1986) with the advent of the flux-corrected transport algorithm (Boris and Book 1973). This scheme for the first time yielded well-behaved solutions of the continuity equation in transition zones

between zero and nonzero layer thickness where under- and overshoots (especially negative Δp values) are particularly detrimental to numerical stability. An alternative approach to pure isentropic modeling, advocated by Hsu and Arakawa (1990), is to compute horizontal mass fluxes using the Takacs (1985) advection scheme. Since this scheme does not enforce positive definiteness as rigorously as flux-corrected transport does, Hsu and Arakawa (1990) had to keep a small amount of mass in coordinate layers that in the aforementioned models are allowed to become truly massless.

Having assembled a set of numerically resilient tools for handling the intersection of isentropes with the ground, the modeling community turned its attention to the second problem in isentropic modeling: the need to accomodate unstable lapse rates associated with diabatic surface forcing. Since both problems manifest themselves at or near the surface, the remedies developed to address the coordinate–ground intersection problem also were useful in alleviating the unstable-lapse-rate problem. Models general enough to accomodate diabatic surface forcing (Zhu et al. 1992; Bleck and Benjamin 1993; Zapotocny et al. 1994; Konor and Arakawa 1997; Johnson and Yuan 1998; Schaak et al. 2004; Benjamin et al. 2004) therefore resemble the ones mentioned earlier in their vertical layout. The salient aspect of all of these approaches is that they "hybridize" the isentropic coordinate (i.e., strike a compromise between Lagrangian and Eulerian vertical grid representation). This is also true for the present model and reflects our conviction that, after more than three decades of isentropic/isopycnic model development, hybridization has emerged as the optimal strategy. In fact, the focus of isentropic model development in the last 20 years has been entirely on hybridization details.

3. Brief review of FIM dynamics

The horizontal aspects of the dynamical core of FIM are described in detail elsewhere (Lee and MacDonald 2009), but a brief review is needed here to put the ensuing discussion of the vertical coordinate into context. Readers unfamiliar with equations in layer and/or generalized vertical coordinate form are referred to Starr (1945), Kasahara (1974), and Bleck (1978a).

FIM solves a system of layer-integrated, hydrostatically approximated conservation equations for momentum, mass, thermal energy, and gaseous as well as liquid or frozen water content. Its so-called physics components that model diabatic forcing of the atmosphere—primarily water phase changes, radiation, vertical turbulent mixing, and surface fluxes—were imported from the Global Forecast System (GFS) of NOAA's National Centers for Environmental Prediction (NCEP). Details about their

implementation will be the topic of a forthcoming paper on FIM's performance in general.

In the following, s is the vertical coordinate, $\mathbf{v}$ is the (Cartesian) horizontal velocity vector taken as vertically constant within an s layer, ∇_s is the two-dimensional gradient operator at s = const, $\Pi = c_p(p/p_0)^{R/c_p}$ is the Exner function, $\theta = c_p T/\Pi$ is the virtual potential temperature, $M = gz + \Pi\theta$ is the Montgomery potential, ζ is the vorticity (i.e., the vertical or $\mathbf{k}$ component of the velocity curl vector), $\dot{\theta}$ is the net diabatic heating, and $\mathbf{F}$ is the sum of frictional forces.

The layer-integrated conservation equations, supplemented by the hydrostatic equation, are given as follows. The momentum conservation equation is

$$\frac{\partial \mathbf{v}}{\partial t} + (\zeta + f)\mathbf{k} \times \mathbf{v} + \frac{1}{\Delta p}\left[\left(\dot{s}\frac{\partial p}{\partial s}\right)_2(\hat{\mathbf{v}}_2 - \mathbf{v}) - \left(\dot{s}\frac{\partial p}{\partial s}\right)_1(\hat{\mathbf{v}}_1 - \mathbf{v})\right]$$
$$+ \nabla_s\left(M + \frac{\mathbf{v}^2}{2}\right) - \Pi\nabla_s\theta = \mathbf{F}. \tag{3}$$

Here, indices 1 and 2 denote the upper and lower interface of a coordinate layer, and $\Delta p = p_2 - p_1$. The vertical advection terms (those involving $\dot{s}$) make reference to interface velocity values denoted here by $\hat{\mathbf{v}}$. Since $\mathbf{v}$ is discontinuous at interfaces, the definition of $\hat{\mathbf{v}}$ is to some extent arbitrary and in practice depends on the chosen vertical advection scheme. The equations for mass conservation and thermal energy conservation are, respectively,

$$\frac{\partial \Delta p}{\partial t} + \nabla_s \cdot (\mathbf{v}\Delta p) + \left(\dot{s}\frac{\partial p}{\partial s}\right)_2 - \left(\dot{s}\frac{\partial p}{\partial s}\right)_1 = 0 \quad \text{and} \tag{4}$$

$$\frac{\partial(\theta\Delta p)}{\partial t} + \nabla_s \cdot (\theta\mathbf{v}\Delta p) + \left(\dot{s}\frac{\partial p}{\partial s}\hat{\theta}\right)_2 - \left(\dot{s}\frac{\partial p}{\partial s}\hat{\theta}\right)_1 = \dot{\theta}\Delta p. \tag{5}$$

As before, the caret in (5) denotes interface values needed in the vertical transport terms. The method by which they are constructed (upstream, centered, ...) determines properties such as monotonicity and diffusiveness of vertical advection in the model. Equations for other mass field tracers (moisture, etc.) have the same form as (5). Last, the hydrostatic equation is

$$\frac{\partial M}{\partial \theta} = \Pi. \tag{6}$$

Given the possibility of large temporal changes in vertical grid spacing, much attention is paid in FIM to conservation of integral properties (i.e., to the elimination of spurious internal sources and sinks for mass and tracer amounts). Solving tracer transport equations in

flux form, as indicated in (5), is one obvious requirement to account for leading-order changes in Δp. (Advective forms are acceptable as long as they are obtained by subtracting from the tracer flux equation the mass continuity equation multiplied by the tracer.) In addition, the various terms in the prognostic equations associated, respectively, with lateral transport, diabatic forcing, vertical grid maintenance, and vertical transport must be evaluated in FIM sequentially for the sake of maintaining a simple and easily understood framework for global conservation.

The specific sequence of operations to advance the model state by one time step, closely following that in the hybrid-coordinate ocean model HYCOM (Bleck 2002), is as follows:

1) The variables u, v, Δp, and q (where q summarily refers to mass field tracers such as θ, humidity, liquid water content, etc.) are updated disregarding all source and vertical transport terms. Layer interfaces are treated as impermeable during this step.
2) Changes to u, v, and q due to model physics are calculated. The physics processes act on the transitional state in which the model is left after completing step 1. The vertical grid remains frozen during step 2.
3) The Δp field resulting from step 1 is modified, where appropriate, by the vertical grid generator described in detail below. We will refer to this as the regridding step.
4) Vertical transport terms necessitated by changes made to Δp in step 3 are evaluated, and their effect is added to the u, v, and q field obtained in step 2. This is the so-called remapping step.

In the special case of zero physical forcing and no action by the grid generator, the transitional model state reached in step 1 becomes the final one.

A few remarks are in order concerning the retrieval of θ from the values $\theta \Delta p$ and Δp obtained in step 1 by solving the horizontal transport portions of (5) and (4). The procedure is potentially ill conditioned in FIM because Δp is not bounded away from zero. To avoid creating nonphysical θ values, we require the ratio $(\theta \Delta p)/\Delta p$ to remain within the range spanned by the old θ values at and near the point in question. While this introduces an element of nonconservation into the transport process, the problem is largely academic because it only occurs where mass flux divergence removes a significant portion, but not all, of the mass from a grid cell. A number of "engineering" fixes are available, such as distributing among neighboring cells the θ amount gained or lost by the clipping process. In present FIM applications, this amount has been found to be too small (in the integrated sense) to warrant corrective action.

4. The ALE coordinate

FIM belongs to a category of circulation models referred to as *layer* models, meaning that vertical spacing of grid points is governed, to varying degrees, by converging or diverging horizontal mass fluxes. Since the prognostic equations resemble the shallow-water equations—even in the sense that variables carried within individual layers are for many intents and purposes treated as vertically constant, layer models are also referred to as stacked shallow-water models.

The hybrid grid in FIM resembles that of the Rapid Update Cycle (RUC; Benjamin et al. 2004; Bleck and Benjamin 1993), but vertical staggering of variables is different because, while RUC conserves mass, it does not rigorously conserve other mass field constituents. In RUC, θ is carried on interfaces as in the Charney–Phillips grid (Arakawa and Moorthi 1988). This staggering convention, which is also followed in the Konor and Arakawa (1997) model, makes restoration of θ to its prescribed coordinate value (see below) somewhat easier in RUC than in FIM, which uses Lorenz-type staggering (Arakawa and Moorthi 1988). FIM staggering replicates the layer treatment in HYCOM (Bleck 2002) in which only pressure and geopotential are carried on interfaces while all other variables, including tracers and Montgomery potential, are defined in layers.

The vertical coordinate in FIM is designed around the idea that coordinate layers conform to isentropic layers except in locations where these intersect the earth's surface. There, layers are locally redefined as terrain-following (σ coordinate) layers. An individual coordinate layer can be isentropic in one geographic region and terrain-following in another.

The hybridization concept employed here and in RUC differs from hybrid schemes developed elsewhere (Bleck 1978b; Pierce et al. 1991; Zapotocny et al. 1991, 1994; Konor and Arakawa 1997; Johnson and Yuan 1998; Webster et al. 1999) in that it relies on locally mandated adjustments of vertical grid spacing rather than on a fixed formula typically consisting of a weighted average of two or more traditional coordinate choices. The vertical coordinate in FIM is best described as arbitrary Lagrangian–Eulerian (Hirt et al. 1974), but we have added a mechanism for keeping coordinate layers aligned with their designated target isentropes over time whenever and wherever this does not conflict with minimum layer-thickness constraints. The original ALE scheme (Hirt et al. 1974) only concerned itself with the maintenance of nonzero grid spacing in Lagrangian coordinate simulations. Coordinate "maintenance" in the sense of keeping coordinate surfaces aligned with isosurfaces of some quasi-conservative physical property is

essential if an ALE-type circulation model is to be used in long-range weather or climate simulation.

While the flexibility of coordinate placement in ALE-type schemes is disconcerting to some users because grid point location in model space cannot be expressed in terms of an analytic formula, the scheme excels in maximizing the size of the purely isentropic subdomain. This follows from the scheme's ability to set the height where the σ coordinate gives way to the θ coordinate in each geographic location separately (i.e., unencumbered by global considerations). Thus, while the lowest purely isentropic coordinate surface in schemes D of Bleck (1978b) and Zapotocny et al. (1994) in a global model must be selected with the highest summertime temperature over the Himalayas in mind, the ALE approach contains no such restrictions. Furthermore, the problem of coordinate surface intertwining due to inappropriately chosen coordinate blending coefficients (Zhu et al. 1992) does not arise in an ALE scheme.

FIM manages the vertical grid structure as follows. If a given layer is "on target" (meaning that θ matches the target potential temperature assigned to this layer) and if, in addition, the two-dimensional shallow-water continuity equation given in (2) [which is (4) without the $\dot{s}$ terms] yields a layer thickness Δp that does not fall below a predetermined minimum $\Delta p_{\min}$, the Δp value obtained from (2) is accepted. In other words, FIM sets $\dot{s} = 0$ in this case, meaning that it treats the interfaces above and below the layer in question as material. If one of the above conditions is not met, the grid generator (see following section) takes over and changes $\partial \Delta p / \partial t$ in a way that either maintains minimum thickness or, if the layer has become separated from its target potential temperature, moves it closer to it. In these situations, the selected Δp tendency is inserted into the full continuity equation given in (4), which at this point becomes a diagnostic equation for the interlayer mass fluxes $\dot{s} \partial p / \partial s$. The latter are used to vertically advect momentum and other variables.

The above process is complicated by the need to include conditions in at least two adjacent layers when deciding on the value of $\dot{s} \partial p / \partial s$ on a given interface. However, as long as $\Delta p_{\min} > 0$ is imposed only on layers at the bottom or top of the column, the test for nonzero $\dot{s} \partial p / \partial s$ can be carried out recursively by a single sweep up or down the column (i.e., it does not require iteration). In this narrow sense, grid maintenance in the present ALE-like scheme appears to be slightly simpler than that described in section 3f of Konor and Arakawa (1997), which requires iterations.

The process described above can be summarized as follows. All hydrostatic models infer the vertical component of motion from the vertically integrated horizontal mass flux divergence. The grid generator in an ALE-type model divides this material vertical motion into two components: vertical motion of the coordinate surface and vertical air motion relative to it:

$$
\begin{pmatrix}
\text{vertical} \\
\text{motion} \\
\textit{of} \\
\text{interface}
\end{pmatrix}
+
\begin{pmatrix}
\text{vertical} \\
\text{motion} \\
\textit{through} \\
\text{interface}
\end{pmatrix}
=
\begin{pmatrix}
\text{vertically} \\
\text{integrated} \\
\text{horizontal} \\
\text{mass-flux} \\
\text{divergence}
\end{pmatrix}. \tag{7}
$$

The decision whether to accept the solution of (2)—that is, whether to balance the right-hand side of (7) by only the first term or by some combination of both terms on the left—is made by the grid generator at each grid point and each time step individually. The FIM grid generator actually performs both this task and the one listed earlier as step 4: it carries out the vertical remapping of all prognostic variables to the modified grid. As noted before, remapping is formally equivalent to vertical advection because it is driven by nonzero values of $\dot{s}$. However, since vertical displacement of atmospheric constituents due to actual air motion is already accounted for in the heaving and slumping of coordinate layers in step 1, vertical advection via the $\dot{s}$ terms in the prognostic equations is best viewed as a secondary property redistribution necessitated by coordinate surfaces migrating through resting air. With the atmosphere conceptually "frozen" in time during this redistribution, remapping should conserve certain integral properties such as column integrals of momentum, thermal energy, and so on.

Turbulent vertical mixing is typically parameterized in large-scale atmospheric models by solving a vertical diffusion equation with an eddy diffusivity coefficient designed to incorporate the effects of stratification, wind shear, and so on. In an isentropic framework, one of the diffused quantities is an independent variable, namely, the vertical coordinate θ. The approach taken in FIM to deal with this peculiarity of isentropic modeling is described in appendix A. The scheme is primarily employed as a safeguard against layer collapse during frontogenesis. One may view it as a rudimentary attempt at qualitatively parameterizing the effect of clear-air turbulence.

5. The vertical grid generator

a. Background

Owing to differences in the vertical staggering of variables, not all schemes existing today for maintaining the vertical grid structure in ALE-type hybrid-isentropic

layer models are interchangeable. The scheme developed for RUC (Bleck and Benjamin 1993; Benjamin et al. 2004) in particular cannot be used directly in FIM. Instead, the technology suitable for the staggered FIM grid had to be imported from HYCOM.

The first-generation HYCOM grid generator, whose design principles are described in detail in Bleck (2002) but date back to Bleck and Boudra (1981), has been modified and tuned over the years to address grid degeneracies that came to light as the range of applications of HYCOM grew. This tuning has added branches to the decision tree in the original algorithm, creating a situation in which the underlying logic is no longer transparent to the user. Complexity in the grid generator discourages experimentation and adaptation of HYCOM/FIM to special modeling tasks, and hence should be avoided.

The algorithm described below represents an attempt to get "back to basics" when moving layer interfaces for the sake of maximizing the part of the atmosphere represented by isentropic layers while at the same time subjecting layers to minimum-thickness constraints. In the first-generation grid generator, each grid point is inspected and adjusted recursively in light of its distance to grid points above and below, using a variety of semi-empirical criteria. The algorithm proposed here is more straightforward in that it begins by transforming a given hybrid stair-step θ profile into a purely isentropic one (i.e., into a stair-step profile whose θ levels are prescribed beforehand (step 1 below). Depending on the stratification and θ range in the original profile, this process can produce massless (zero thickness) layers at the top and bottom of the column. Massless layers that occur at the ground are subsequently inflated to a prescribed minimum thickness (step 2 below).

Differences between the resulting hybridized layer interface pressures and those of the input profile imply mass exchange among layers. Tracers and momentum must then be exchanged between layers as well (step 3 below). Any one of the standard conservative advection schemes can be used for this task.

b. Step 1: Transformation of nonisentropic stair-step θ profiles to isentropic coordinates

Let $\theta_{in}(p)$ be a piecewise-constant ("stair step") vertical profile of θ. Both the step width $\Delta\theta_{in}$ and the step height Δp can vary from step to step. Our task is to transform $\theta_{in}(p)$ into another stair-step profile differing from the original one in that the location of the "risers" on the θ axis is prescribed. Ideally, the transformation should be accomplished without perturbing the potential/internal energy of the column. Another quantity worth preserving is the geopotential height of the column, because a transformation that changes the column height is

likely to set off external gravity waves. In the following we will adopt a height conservation constraint, which is done by switching from p to Π as vertical coordinate (see section 5e below).

Let θ_k ($k = 1, \ldots, n$, $\theta_{k+1} > \theta_k$, k increasing upward) mark the points on the θ axis where we want the new risers to be placed. These points represent the desired coordinate values for the isentropic model domain while $\theta_{in}(\Pi)$ represents the model state at the completion of step 2 above. We require that the θ_k values span the θ range of the input profile:

$$\theta_1 \leq \theta_{in}(\Pi) \leq \theta_n \quad \text{for all } \Pi \tag{8}$$

and that the input profile be monotonic.

As shown in appendix D of Bleck (2002), the interfaces in the transformed profile are given by

$$\Pi_{k+1/2} = \frac{1}{\theta_{k+1} - \theta_k} \int_{\theta_k}^{\theta_{k+1}} \Pi_{in}(\theta) \, d\theta \ (k = 1, \ldots, n-1), \tag{9}$$

where $\Pi_{in}(\theta)$ is the inverse of $\theta_{in}(\Pi)$.

If condition (8) is violated, evaluation of (9) is postponed until the offending input layer is brought into compliance by "diluting" it with mass from adjacent layers. Persistent heating at the model top, for example, is thereby transformed into a gradual thickening of the uppermost coordinate layer.

The transformation of an arbitrary profile $\theta_{in}(\Pi)$ into a stair-step curve with risers at prescribed values θ_k is illustrated in Fig. 1. To make a clear visual impression of the grid generator's action, we have chosen an input profile (thin solid curve) that bears little resemblance to the output profile (dashed). Such large differences only occur during model initialization when the grid generator is employed to transform GFS temperature profiles into hybrid-isentropic space.

FIM initialization is outside the scope of the present article. Nevertheless, an input profile chosen from a set of actual GFS-supplied input profiles provides a welcome opportunity to illustrate not only the action of the grid generator but also an important step in the model initialization procedure. Note that the 10-K spacing of the θ_k values in Fig. 1 is far coarser than the resolution actually used in FIM.

c. Step 2: Enforcement of layer-thickness constraints

Suppose the prescribed potential temperature values θ_1, θ_2, $\ldots$, in the output profile cover a wide enough range to yield $\theta_k < \theta_{in}$ for some $k > 1$. In this case, (9) will yield $\Pi_{1/2} = \Pi_{3/2} = \cdots = \Pi_{k+1/2}$ (i.e., layers $1, \ldots, k$ in the transformed profile will be massless). Likewise, if
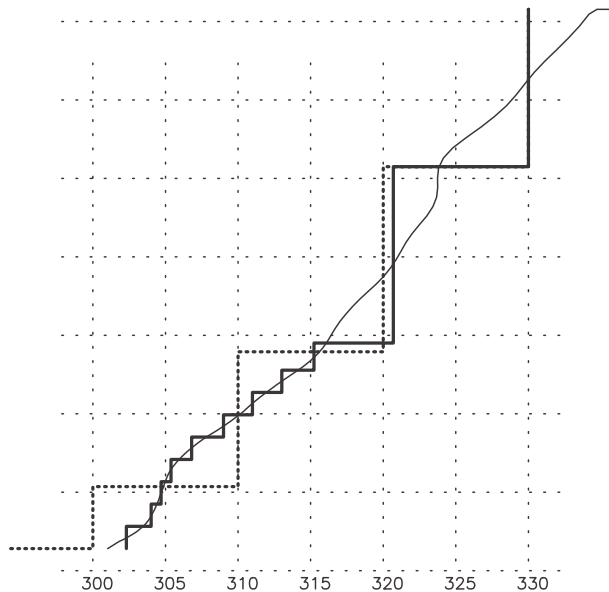
FIG. 1. Example of vertical grid generation in FIM. Potential temperature (K) is on the abscissa and Exner function is on the ordinate. The thin solid line is the input profile $\theta_{in}(\Pi)$, drawn here as a continuous rather than stair-step curve for legibility. The dashed line is the transformed, purely isentropic profile with θ_k values spaced 10 K apart. The thick solid curve is the same profile after inflation of massless bottom layers.

$\theta_{in} < \theta_k$ for some $k < n$, layers $k + 1, \ldots, n$ will be rendered massless $(\Pi_{k+1/2} = \cdots = \Pi_{n+1/2})$.

The strategy in FIM is to accept massless layers aloft but to always inflate massless layers at the bottom of the grid column. Layer inflation rules can be as simple as specifying a constant minimum thickness $\Delta\Pi_0$. In this case the set of isentropic interface values $\Pi_{k+1/2}$ obtained from (9), to be identified here as $\hat{\Pi}_{k+1/2}$ to distinguish them from the final "hybridized" values, are recursively subjected to the constraint

$$\Pi_{k+1/2} = \min(\hat{\Pi}_{k+1/2}, \Pi_{k-1/2} - \Delta\Pi_0) \quad (k = 1, 2, \ldots).$$
(10)

Note that $\Delta\Pi_0$ can easily be made layer- or latitude-dependent or can be scaled by terrain height. Additional refinements of the hybridization scheme are described in appendix B.

d. Step 3: Vertical advection

The regridding process described above must be followed by a remapping step in which model variables are advected vertically in response to changes in interface pressure. Borrowing from HYCOM, vertical advection of momentum and tracers is currently handled by either the piecewise linear method (PLM; van Leer 1974) or the piecewise parabolic method (PPM; Colella and Woodward 1984). Because of the potential presence of massless or near-massless layers, fluid is permitted to

cross more than one interface during a given time step. All variables mentioned are remapped in p space to conserve their mass-weighted column integral.

Potential temperature is a special case. The regridding process described earlier yields a new θ distribution that may be viewed as resulting from upstream or piecewise constant advection in s space. To suppress the numerical diffusivity implied by this low-order scheme, FIM actually discards the θ field resulting from the regridding exercise and replaces it by a field vertically advected by the same higher-order scheme that is used for the other prognostic variables.

There is a price to be paid for inferring the amount of mass transferred between layers from a piecewise constant θ distribution, as is done in (9), and subsequently using a higher-order scheme to remap θ. Neither will the slab of air arriving in a layer have the potential temperature needed to precisely restore that layer to target, nor will the transfer leave θ in the donor layer unchanged. However, we find that the restoring algorithm allows layers to lock onto their targets relatively quickly.

It is worth noting that the conservation properties of the regridding/remapping algorithm are determined during the remapping phase alone. Hence, regridding can be based on Π, as shown in (9), or on any other function monotonic in pressure. The vertical coordinate used during remapping is also at the discretion of the user. FIM presently offers a choice between $(p/p_0)^\kappa$ and $(p/p_0)^{1+\kappa}$ (where $\kappa = R/c_p$) for remapping the θ field. The rationale for providing these two options is given in the following section.

In Fig. 1, the process of remapping the input profile (thin line) onto the new layer configuration including several inflated bottom layers yields the heavy black curve. This is the final outcome of the three-step procedure.

e. Conservation alternatives

It follows from (6) that the height of an air column can be preserved during vertical remapping by using the Exner function Π as vertical coordinate. Unfortunately, this choice does not allow us to satisfy another important constraint: conservation of column-integrated internal energy $I = \int c_v T \, dp$ and column-integrated potential energy $P = \int g\rho z \, dz$. (In an ideal gas in hydrostatic balance, internal and potential energy are proportional to one another, so conservation of one entails conservation of the other.) The incompatibility of column height conservation with internal/potential energy conservation becomes clear if one writes P and I in terms of θ and p and compares the resulting expressions

$$I = \frac{c_v p_0}{g(1+k)} \int \theta \, d\left(\frac{p}{p_0}\right)^{1+k} \quad \text{and} \quad$$
(11)

$$P = \frac{R p_0}{g(1+k)} \int \theta \, d\left(\frac{p}{p_0}\right)^{1+k} \qquad (12)$$

($k = R/c_p$) with the formula for column height, $\int \theta \, d\Pi$, in which θ is integrated over a variable proportional to $(p/p_0)^{\kappa}$. It is easy to see now that conservation of I and P can be achieved during remapping of θ by using $(p/p_0)^{1+k}$ as vertical coordinate but that this can only be done at the price of violating the height preservation constraint. The relative importance of height versus internal/potential energy conservation is hard to assess without practical tests.

Other options exist. Lin (2004), in his intermittently Lagrangian vertical coordinate scheme, chooses to conserve total energy during regridding and to treat layer temperature as a diagnostic quantity. The reason for giving priority to total energy conservation, as opposed to internal/potential energy conservation, is that dissipation of kinetic energy acts as a heat source. Whether vertical regridding is the proper vehicle for modeling such dissipative processes is a topic inviting further discussion.

f. An illustration

Some design options suitable for an ALE coordinate in atmospheric models are illustrated in Fig. 2. The figure shows a meridional cross section (latitude increasing to the left) that cuts across a typical midlatitude jet stream in thermal wind balance with sloping tropospheric isentropes. Three elements are combined in the figure. Solid lines running across each panel represent layer interfaces. Shaded contours represent zonal wind speed. Colors filling alternate spaces between isotachs show potential temperature. The purpose of the coloration is to indicate where in the domain a given coordinate layer is isentropic. The rendering is not exact because the columnwise-steppy θ field has been converted into a continuous field and interpolated to p space to simplify plotting.

The top-left panel in Fig. 2 shows a layer configuration typically seen in pure isentropic coordinate models like those of Bleck (1984) and Zhu et al. (1992). Since FIM evaluates lateral flux terms in (4) and (5) using the flux-corrected transport scheme, which permits layer thickness to go to zero, FIM can actually operate stably in the pure isentropic mode depicted in that panel. The rationale for building this capability into FIM was to remove numerical-stability related constraints on the choice of minimum layer thickness.

Passing the configuration shown in the top-left panel of Fig. 2 to the FIM grid generator yields the configuration shown in the top-right panel. As described in detail above, the grid generator inflates layers that intersect the ground—the ones shown as massless layers in the top-left panel—but leaves higher layers unmodified. The shallow layers formed in this way near the earth's surface are, of course, no longer isentropic. A configuration similar to this one but with much higher vertical resolution is used in RUC (Benjamin et al. 2004) and in FIM. For illustrative purposes, minimum layer thickness $\Delta p_{\min}$ in Fig. 2 is set to 30 hPa. FIM and RUC typically use values in the range of 2–20-hPa.

The bottom-left panel of Fig. 2 illustrates how the ALE coordinate reacts to the presence of a mountainous feature like the Tibetan Plateau. To keep the mountain from creating a nonisentropic coordinate cap extending to jet stream levels, $\Delta p_{\min}$ is reduced over high terrain as it would be in a σ-coordinate model.

In the bottom-right panel of Fig. 2 we give an example of how one could modify the ALE coordinate to optimize layer spacing at low latitudes where cloud physics parameterization schemes typically demand more uniform vertical grid spacing than that shown in the top-right panel. A layer expansion feature like the one in the bottom-right panel would be particularly desirable in a model that uses an extremely low $\Delta p_{\min}$ for the sake of maximizing isentropic grid representation in the extratropics. Given the large value of 20 hPa currently used by FIM in all but the lowest σ layers, refinements of this type are not contemplated at this time.

Because of the continuous rendering of the steppy θ field, the color fields in Fig. 2 do not accurately reflect θ in each layer, as already mentioned. However, some color irregularities in regions where interfaces have been pushed up or down by the grid generator are the result of vertical advection errors.

This brings up an important point. The flexibility of the ALE scheme might compel a model architect to design a coordinate system that requires interfaces to periodically move over large distances through the fluid. Such a design, an obvious violation of the flow-following concept, can lead to vertical advection and associated dissipation errors that are far in excess of those typically encountered in a fixed grid. One example of an ill-advised ALE algorithm is to attach a coordinate surface to the top of the planetary boundary layer for the sake of cleanly separating well-mixed from stratified air layers. Because of the potentially large difference between day- and nighttime boundary layer heights and the ensuing need to move large amounts of air back and forth across the pulsating interface, advantages gained by separating regions of large and small vertical mixing are likely to be lost. The best strategy is to make coordinate surfaces either truly Lagrangian or, if that is impractical, to keep them fixed in space.

There are situations in which not only time-dependent but also space-dependent $\Delta p_{\min}$ values of the kind
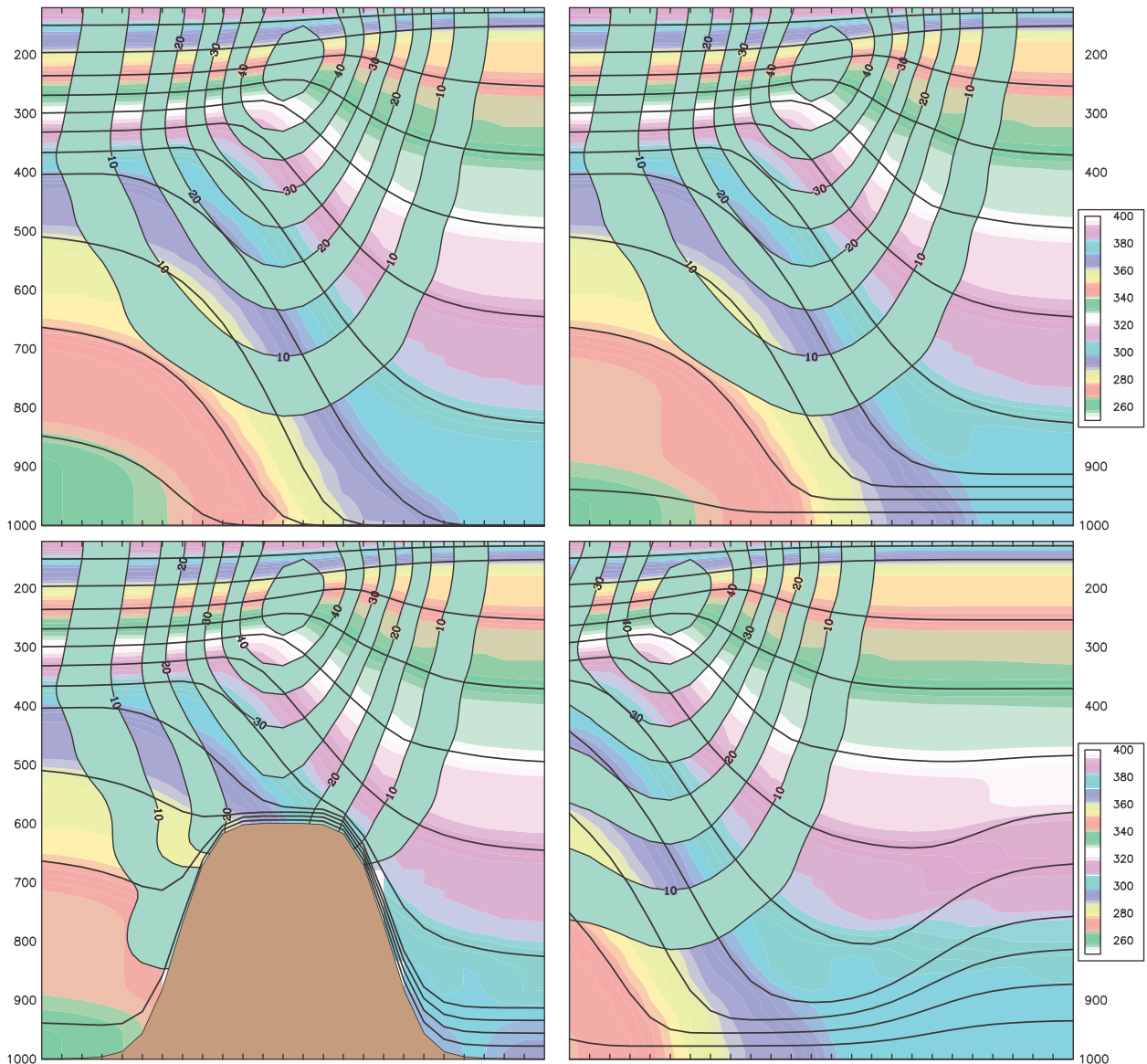
FIG. 2. Vertical–meridional cross section illustrating the functionality of the ALE-like coordinate in FIM. Solid lines are layer interfaces, and shaded contours are isotachs (m s^{-1}). Colors indicate potential temperature (K). The ordinate is pressure (hPa). The horizontal extent is 5000 km.

shown in the top-right panel of Fig. 2 must be avoided. Consider, for example, a model in which the Sadourny (1975) or Arakawa and Lamb (1981) approach is used to formulate the momentum equations in potential enstrophy–conserving form, with layer thickness Δp taking the place of $\partial p/\partial \theta$ in the potential vorticity expression $(\zeta + f)(\partial p/\partial \theta)^{-1}$. In the σ coordinate subdomain, lateral variations imposed on Δp by the grid generator will, in this case, introduce bogus lateral potential vorticity gradients that can affect the flow evolution inappropriately. FIM avoids this pitfall by using vorticity, not potential vorticity, in (3). HYCOM uses

the Sadourny (1975) formulation but makes $\Delta p_{\min}$ a function of the layer index only, effectively replacing potential vorticity by absolute vorticity inside the p coordinate subdomain.

6. A sensitivity study

Because the focus of this article is on the design of an ALE-type vertical coordinate, readers might expect to see results of experiments highlighting the performance of the grid generator. While such experiments have been carried out, and forecast quality has indeed been found

to depend on parameters such as minimum layer thickness, a scientifically meaningful discussion of these experiments will require exploring the fundamental question of how sensitive today's cloud physics and boundary layer parameterizations are to deviations from the vertical resolution provided by conventional NWP models. This question, though eminently important, is beyond the scope of the present article.

One option would have been to present results gained with a FIM version stripped entirely of its physics components. However, we argue that the results presented below, which highlight FIM's capabilities as a full-fledged global medium-range prediction model, are more informative than process studies dwelling on coordinate configurations in idealized adiabatic flow, especially since such experiments, which naturally were carried out during the early stages of FIM development, did not reveal ALE-related numerical issues.

In light of the above, the material presented below dwells on capabilities of FIM associated with its isentropic coordinate, rather than on the particulars of low-level σ-layer packing or the σ–θ transition.

A displaced fluid parcel that does not experience a buoyancy force driving it back to its original location is likely to remain in contact with its new surroundings longer than it otherwise would be. This gives the parcel some extra time to exchange properties with the surrounding fluid (Montgomery and Spilhaus 1941, p. 281). Consequently, turbulent mixing in stratified fluids takes place preferentially along surfaces of constant potential buoyancy (i.e., buoyancy corrected for compressibility effects).

With potential buoyancy and entropy being synonymous (or nearly so) in the atmosphere, turbulent exchange tends to minimize isentropic gradients of properties such as momentum (Rossby and Collaborators 1937). Replicating this process in a numerical model is not easy if coordinate surfaces do not coincide with isentropes. Dispersion errors associated with horizontal transport in a model tend, over time, to destroy property contrasts on whichever surfaces the transport is being carried out. Only if the transport equations are solved on isentropic surfaces can this numerical dispersion error be hidden behind what we may call a "smoke screen" of naturally occurring mixing. Being able to do this is especially advantageous in long-term simulations of statistically stationary or slowly varying states that in nature result from a balance among external forcing, transport, and turbulent mixing.

Note that the above argument refers to numerical errors associated with *lateral transport*—that is, errors caused by the dispersive properties of the horizontal advection operator. Errors arising during evaluation of explicit mixing terms (of which FIM is free) can also be important. However, these can be reduced—to some extent at least—by aligning the main axes of the mixing tensor with isentropic surfaces (Redi 1982). An equivalent strategy to project the effect of numerical dispersion in the transported field onto isentropic surfaces has not yet been developed.

The flexibility of the ALE coordinate allows us to shed light on the correctness of the assertion that numerical accuracy of transport processes benefits from isentropic coordinate representation. We will do this by changing the vertical coordinate in FIM from hybrid σ–θ to a more traditional combination of σ and p and will look for at least anecdotal evidence that simulations based on the first-mentioned coordinate yield more coherent patterns of dynamically relevant quantities than simulations based on the latter. This work is complementary to earlier work by, among others, Johnson et al. (1993) and Benjamin et al. (2004), but sheds light on fluid dynamics phenomena that are not specifically treated in those studies.

A tracer well suited for this purpose is the vorticity $\zeta = \partial v/\partial x - \partial u/\partial y$. Even though ζ is neither explicitly advected in primitive equation models nor rigorously conserved, the fact that vorticity is composed of spatial derivatives of the velocity field and interacts with the circulation in a two-way mode makes it a particularly sensitive indicator of forecast errors. We will focus in the following on the process of tropospheric Rossby wave breaking (McIntyre and Palmer 1985), also referred to as vortex rollup (Dritschel and Polvani 1992) or, in synoptic meteorologists' parlance, cutoff-low formation.

Three synoptic cases were analyzed in detail for this article. Space limitations permit us to present only one of them in detail. Limited results from the other cases will be included to indicate that trends in forecast accuracy as functions of horizontal resolution and vertical grid were not unique to the first case.

The vortex rollup process and the failure of some model versions to simulate it accurately were found to be depicted most succinctly by the vorticity and geopotential height distribution on the 300-hPa isobaric surface. We will therefore confine our attention to 300-hPa flow patterns and will first focus on forecasts over North America extending 3.5–5 days from the initial time of 0000 UTC 19 October 2008. (The other two cases depict vortex rollup events over Europe 4 weeks later and over the Southern Ocean in late austral summer.)

Eight FIM forecasts were carried out, four using the native FIM coordinate and four using the σ–p coordinate introduced to the operational GFS in May 2007. The switch from the former coordinate to the latter is accomplished in FIM by replacing the standard hybrid-isentropic grid generator by one that simply restores interface pressures to the values used in the GFS. No other changes are made to the model.

The forecasts within each group of four differ by horizontal grid resolution. As outlined in section 3a of Lee and MacDonald (2009), recursively bisecting the sides of the 20 triangles in the icosahedron quadruples the number of hexagonal cells on the sphere. In the experiments reported here, the number of cells ranges from ~10 000 (referred to as G5 resolution: 5 bisecting steps) to ~655 000 (G8: 8 bisecting steps). The mesh size is ~240 km at G5 resolution and ~30 km at G8 resolution. Since icosahedral surface elements have to be projected outward onto the enclosing sphere, mesh size varies by approximately 15%.

The figures for case 1 are organized as follows. Results obtained by FIM configured with its standard hybrid $\sigma-\theta$ coordinate are displayed in the top four panels of Figs. 3–6. Corresponding results obtained by substituting the hybrid $\sigma-p$ coordinate for the native FIM coordinate are shown in the bottom four panels of each figure. As already mentioned, coordinate values in the $\sigma-p$ grid are identical to those used by the GFS. Both grids consist of 64 layers. The noisy height field seen in some figures is partly due to deficiencies in vertical and horizontal interpolation.

We begin by showing in Fig. 3 the rollup process as simulated at the highest available resolution of 30 km (G8). The forecasts clearly depict the universal process by which mixing in fluids takes place: initally compact fluid elements are continually stretched into long, thin filaments that create sharp property gradients subsequently eroded by molecular diffusion. How important explicit simulation of this stretching or stirring process is for climate modeling is an important open question. There is no doubt, however, that such explicit simulation is important in NWP because property gradients formed by filamentation often spawn severe local weather events.

Figure 3 indicates that the $\sigma-\theta$ forecast carries the filamentation process somewhat farther than does the $\sigma-p$ forecast. In fact, it may carry it too far. A comparison with observed conditions on 23 October 2008 (not shown here) indicates that FIM, using its native coordinate, actually overintensifies the cutoff vortex. Further experimentation will be required to determine whether omission of explicit subgrid-scale mixing terms in FIM is an occasional detriment to forecast accuracy.

The next figures depict the rate at which the filamentation and rollup process is degraded with lower grid resolution. At G7 (60 km) resolution (Fig. 4), the vorticity streamer in the $\sigma-\theta$ forecast is seen to maintain its integrity, even to the extent that its spiral structure at 120 h is better defined in this forecast than in the previous one. Larger changes between Fig. 3 and Fig. 4 are noticeable in the $\sigma-p$ forecast, and these changes foreshadow a precipitous decline in filament definition with decreasing grid resolution in $\sigma-p$ mode. The 300-hPa

surface at 120 h in the $\sigma-p$ forecast has risen by 10 m as a result of the resolution change, and the vortex is located too far to the west.

It is worth noting that even in a perfect model simulation, the filament patterns in $\sigma-\theta$ and $\sigma-p$ forecasts would differ because the plotted vorticity is based on winds differentiated at constant θ in one model version and at constant p in the other. (At 300 hPa, the GFS coordinate is nearly isobaric while FIM's hybrid-isentropic coordinate is solidly isentropic.)

The trend suggested by Figs. 3 and 4 continues as mesh size is doubled again, to 120 km (Fig. 5). At this resolution, the vorticity streamer in the $\sigma-\theta$ forecast loses its distinctive spiral character, but its forward edge still shows signs of being wrapped around the vortex at 120 h. In the $\sigma-p$ forecast at 120-km resolution, the vorticity field is devoid of sharp maxima and the rollup process is greatly weakened. The trough line develops a serious tilt toward the southwest.

At 240-km spatial resolution (Fig. 6) the $\sigma-\theta$ forecast finally shows signs of serious degradation, reminiscent of what we saw at higher resolution in the $\sigma-p$ forecast. Vortex rollup is no longer taking place, even though the erroneous tilt of the trough line is less serious in Fig. 6 (top panels) than in the 120-km $\sigma-p$ forecast (bottom panels of Fig. 5). The trough in the 240-km $\sigma-p$ forecast no longer shows signs of amplification during the 84–120-h time frame.

As previously mentioned, because of space limitations and to avoid being unduly repetitive we will present material from the other two cases only as needed to amplify the points made above.

G8 forecasts over central Europe extending 5.5 and 6 days out from the starting date of 0000 UTC 17 November 2008 (Fig. 7) show the by now familiar differences between the two FIM versions. The vorticity streamers in the $\sigma-\theta$ forecast look tighter, and while the speed at which they wrap around the vortex appears to be similar in the two forecasts, the total amount of vorticity drawn into the vortex—if this can be judged by looking at a map—appears to be stronger in the $\sigma-\theta$ forecast, explaining the extra deepening at 144 h relative to that in the $\sigma-p$ forecast.

Like in the earlier case, the discrepancy between $\sigma-p$ and $\sigma-\theta$ forecasts becomes more pronounced with lower horizontal resolution. We only show results here for the G7 (60-km mesh size) experiments (Fig. 8). Streamers have virtually disappeared from the $\sigma-p$ forecast, and there is only a hint of a cutoff low. The $\sigma-\theta$ forecast fares better. In fact, there is a striking resemblance between the vorticity patterns in the G7 $\sigma-\theta$ forecast and the G8 $\sigma-p$ forecast.

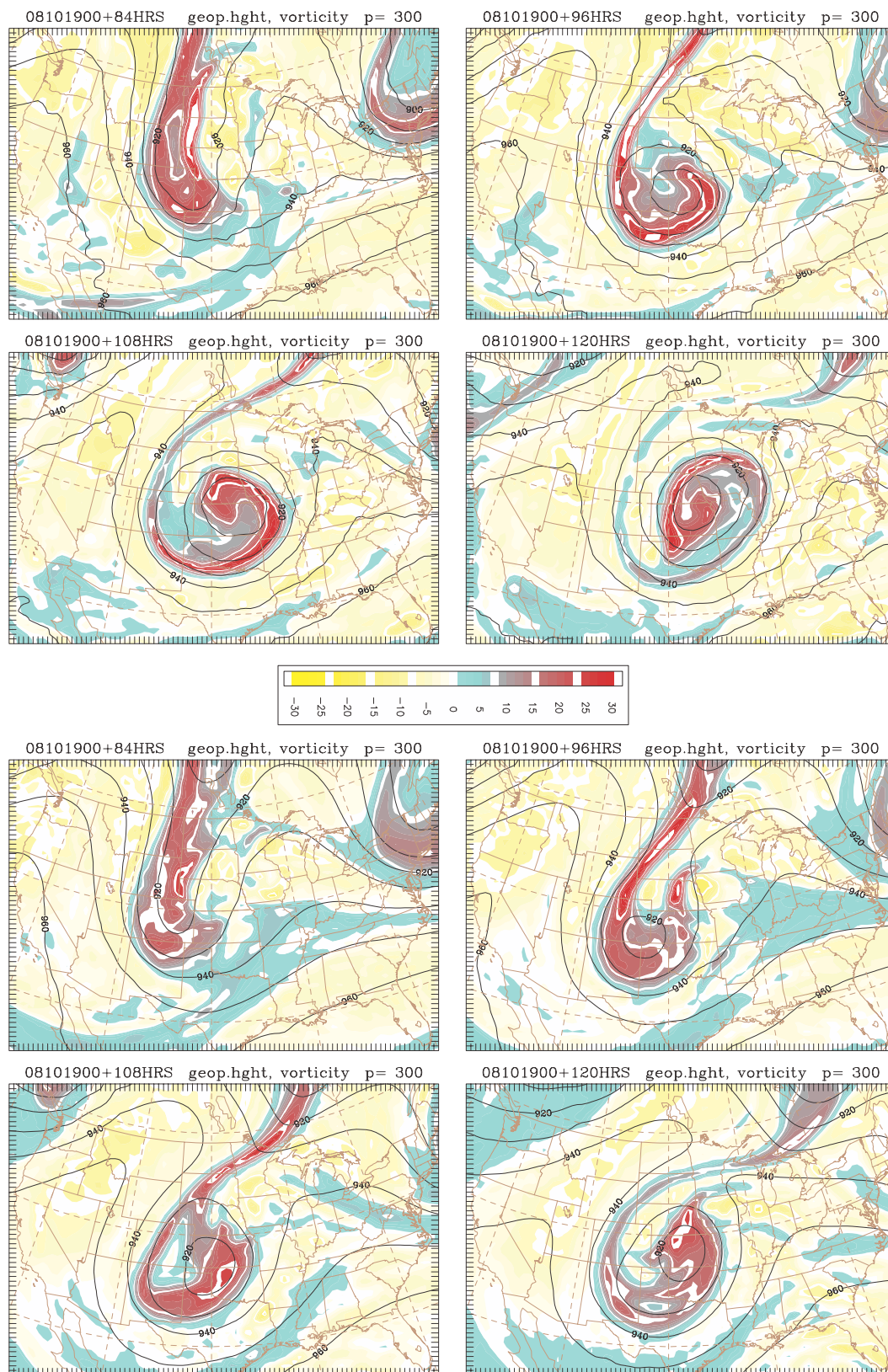Three-day and 3.5-day 300-hPa forecasts at G8 resolution over the Indian Ocean sector of the Southern

08101900+84HRS geop.hght, vorticity p= 300

08101900+96HRS geop.hght, vorticity p= 300

08101900+108HRS geop.hght, vorticity p= 300

08101900+120HRS geop.hght, vorticity p= 300

08101900+84HRS geop.hght, vorticity p= 300

08101900+96HRS geop.hght, vorticity p= 300

08101900+108HRS geop.hght, vorticity p= 300

08101900+120HRS geop.hght, vorticity p= 300

FIG. 3. 84-, 96-, 108-, and 120-h forecasts of geopotential height (m) and vorticity (10^{-5} s^{-1}) at 300 hPa. The initial time is 0000 UTC 19 Oct 2008. The top four panels are forecasts based on the native σ–θ grid, and the bottom four panels are FIM forecasts using the GFS σ–p grid. The horizontal resolution is G8 (~30 km).

08101900+84HRS geop.hght, vorticity p= 300

08101900+96HRS geop.hght, vorticity p= 300

08101900+108HRS geop.hght, vorticity p= 300

08101900+120HRS geop.hght, vorticity p= 300

08101900+84HRS geop.hght, vorticity p= 300

08101900+96HRS geop.hght, vorticity p= 300

08101900+108HRS geop.hght, vorticity p= 300

08101900+120HRS geop.hght, vorticity p= 300
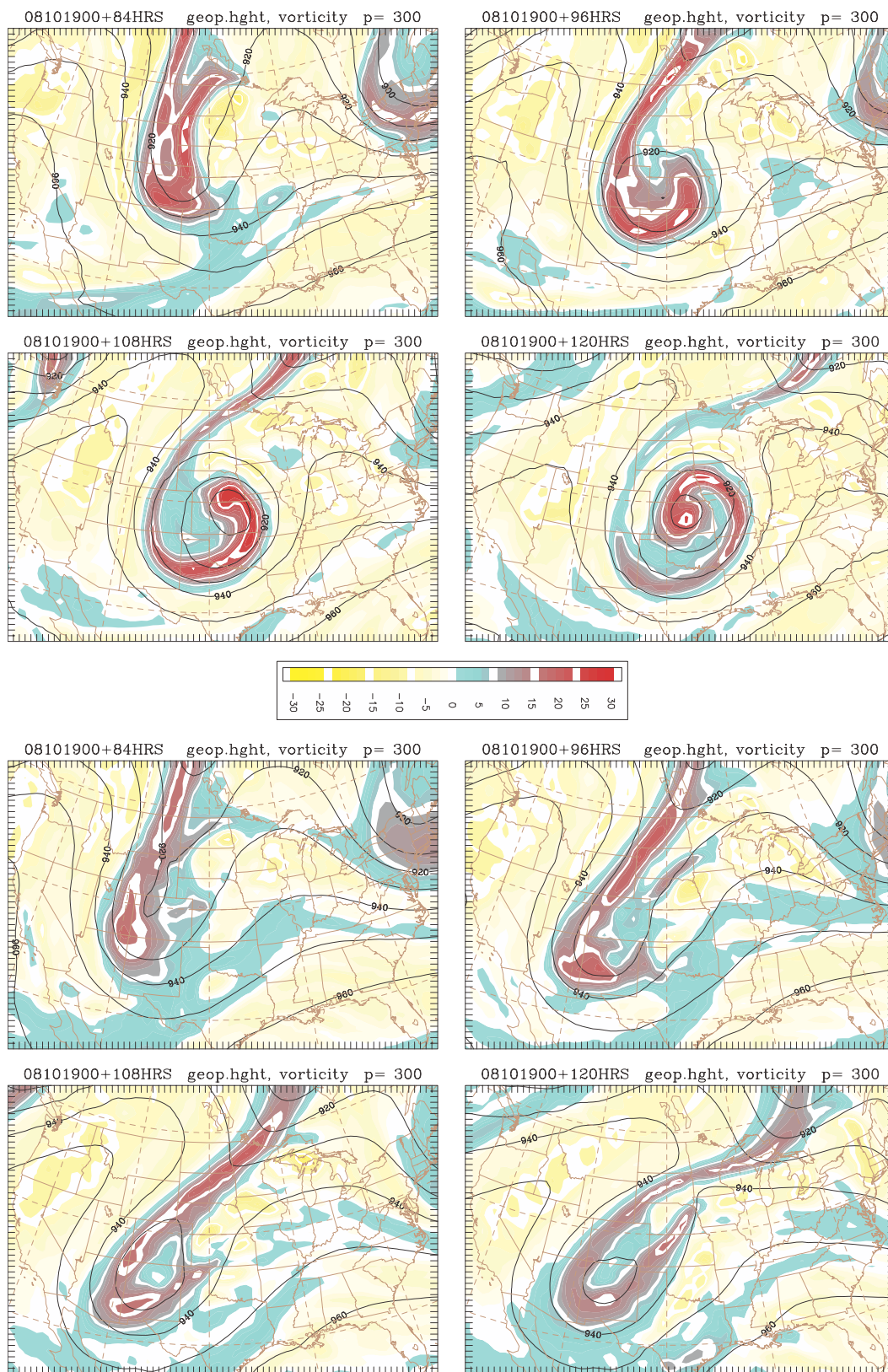
FIG. 4. As in Fig. 3, but for G7 (~60 km) resolution.

08101900+84HRS   geop.hght, vorticity   p= 300

08101900+96HRS   geop.hght, vorticity   p= 300

08101900+108HRS   geop.hght, vorticity   p= 300

08101900+120HRS   geop.hght, vorticity   p= 300

08101900+84HRS   geop.hght, vorticity   p= 300

08101900+96HRS   geop.hght, vorticity   p= 300

08101900+108HRS   geop.hght, vorticity   p= 300

08101900+120HRS   geop.hght, vorticity   p= 300
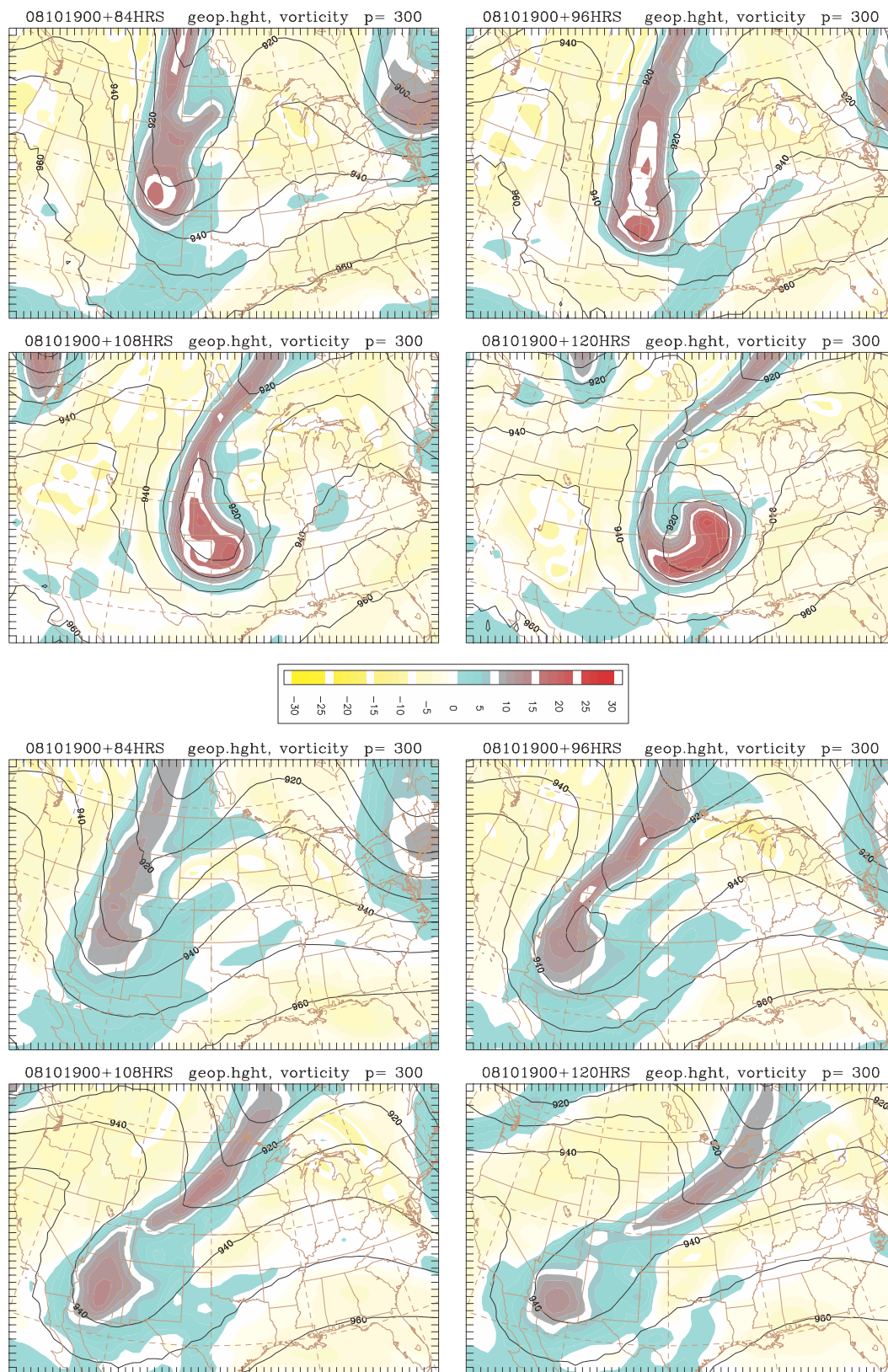
FIG. 5. As in Fig. 3, but for G6 (~120 km) resolution.

08101900+84HRS   geop.hght, vorticity   p= 300

08101900+96HRS   geop.hght, vorticity   p= 300

08101900+108HRS   geop.hght, vorticity   p= 300

08101900+120HRS   geop.hght, vorticity   p= 300

08101900+84HRS   geop.hght, vorticity   p= 300

08101900+96HRS   geop.hght, vorticity   p= 300

08101900+108HRS   geop.hght, vorticity   p= 300

08101900+120HRS   geop.hght, vorticity   p= 300
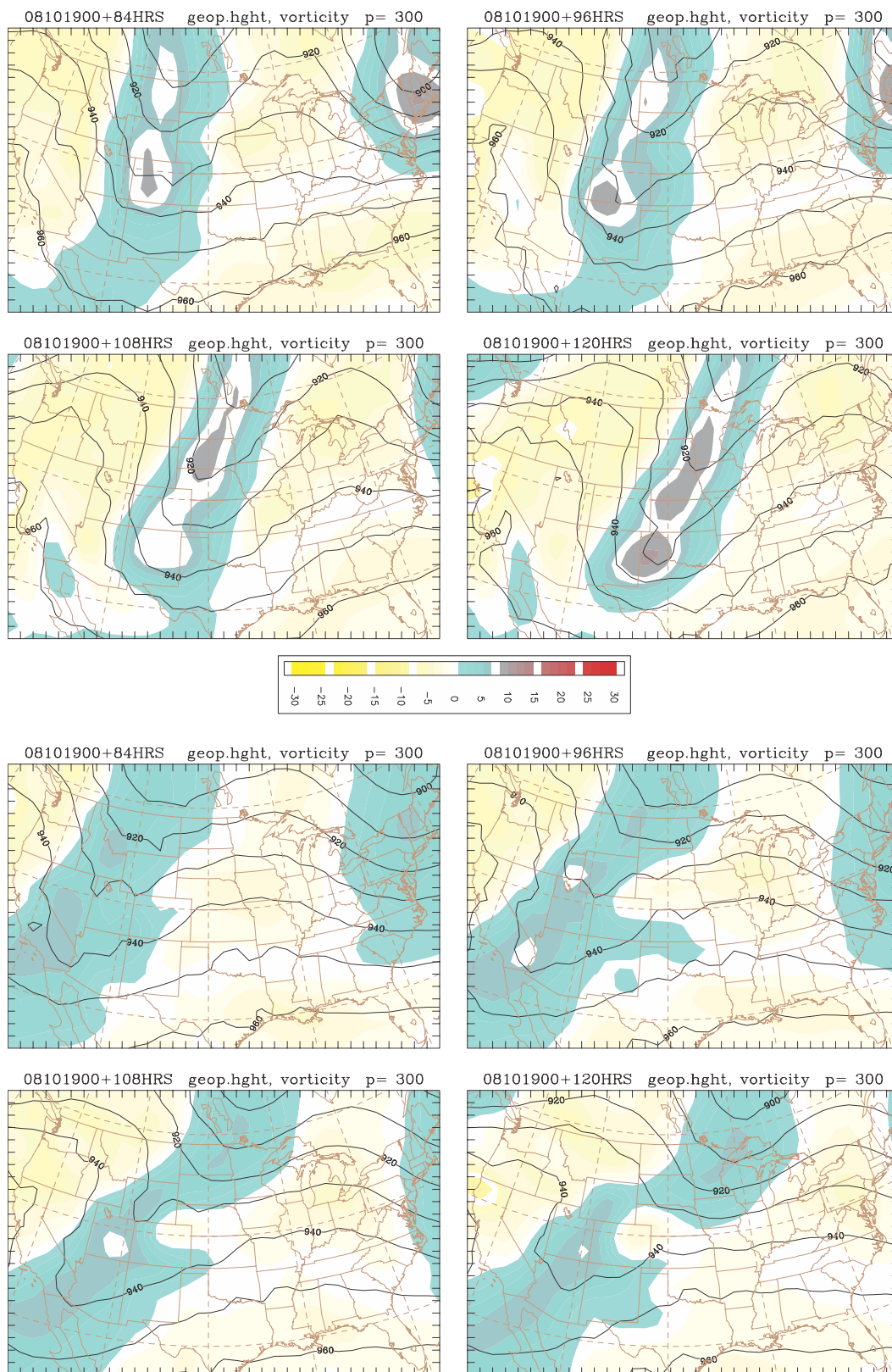
FIG. 6. As in Fig. 3, but for G5 (~240 km) resolution.

08111700+132HRS   geop.hght, vorticity   p= 300   08111700+144HRS   geop.hght, vorticity   p= 300

08111700+132HRS   geop.hght, vorticity   p= 300   08111700+144HRS   geop.hght, vorticity   p= 300
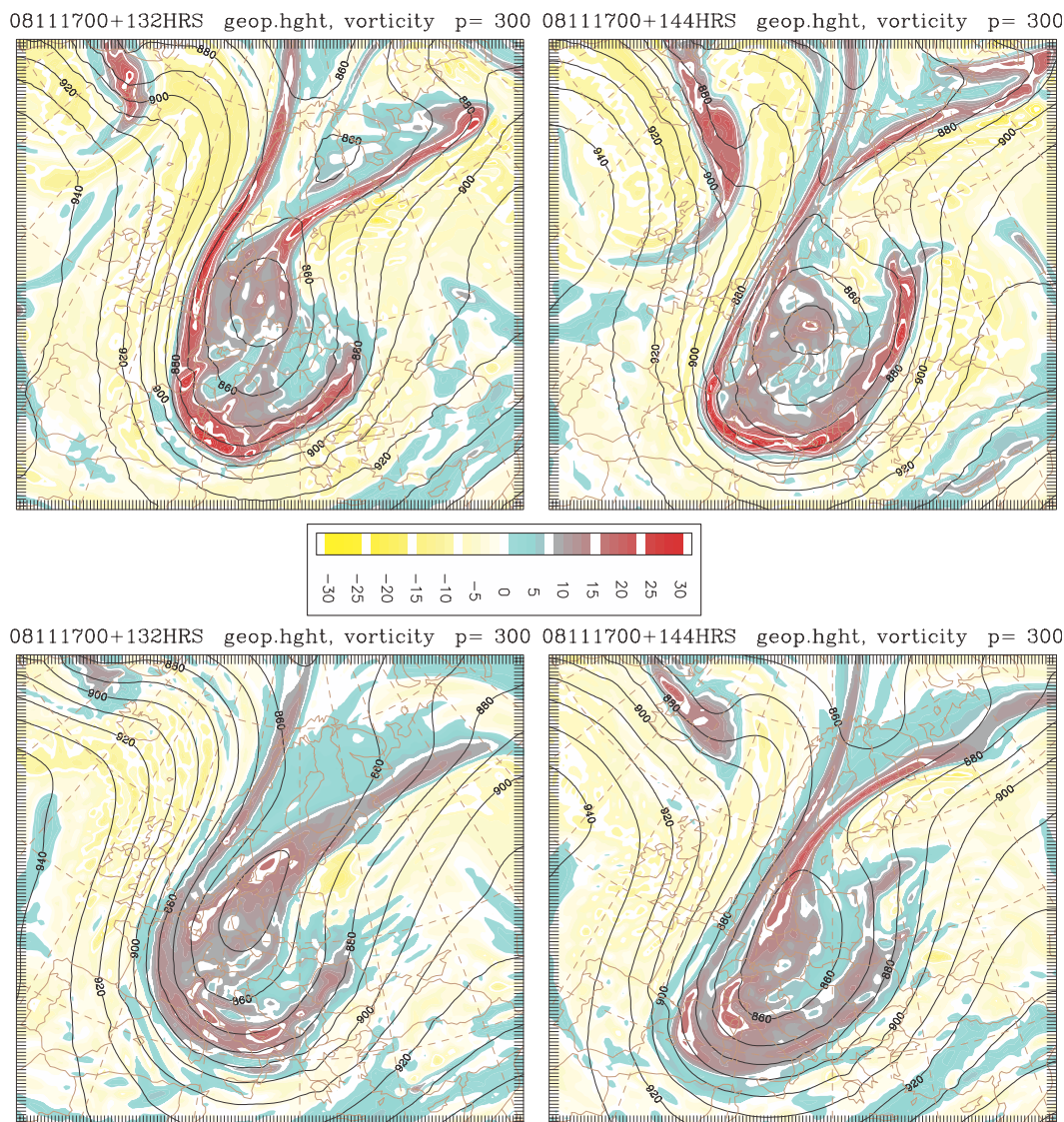
FIG. 7. 132- and 144-h forecasts of geopotential height (m) and vorticity (10^{-5} s^{-1}) at 300 hPa. The initial time is 0000 UTC 17 Nov 2008. (top) Forecasts based on native σ–θ grid. (bottom) FIM forecasts using GFS σ–p grid. The horizontal resolution is G8 (~30 km).

Ocean are shown in Fig. 9. The initial time is 0000 UTC 7 February 2009. This is a late summer case; hence, vorticity contrasts are not as strong as in the other cases. The color scale has been adjusted to take this into account.

Differences in the large-scale vorticity pattern between σ–p and σ–θ forecasts are minor at G8 resolution, even though the vorticity filaments appear to be slightly more coherent in the σ–θ forecast.

The weakening of the wave-breaking process with decreasing horizontal resolution is not as pronounced as in the previous cases, but the by-now familiar pattern of degradation can still be detected. Despite the relatively low resolution of 120 km in the G6 forecasts shown in Fig. 10, the remnants of a vorticity spiral are

still very noticeable at 72 h in the σ–θ forecast (top-left panel) whereas there is nothing left of it in the σ–p forecast. As in the November 2008 case, the "amount" of vorticity trapped in the vortex appears to be higher in the σ–θ forecast, leading to a slightly stronger cutoff vortex.

An alternate view of the difference between the two forecasts is provided in Fig. 11. The cross sections shown correspond in time and grid resolution to the fields in Fig. 10. The sections slice through the vortex in zonal direction, thereby creating a butterfly pattern in the velocity field. The jet in the σ–θ forecast is the stronger one, by roughly 5 m s^{-1} on the upstream side and 10 m s^{-1} on the downstream side. The difference in strength is

08111700+132HRS geop.hght, vorticity p= 300 08111700+144HRS geop.hght, vorticity p= 300

08111700+132HRS geop.hght, vorticity p= 300 08111700+144HRS geop.hght, vorticity p= 300
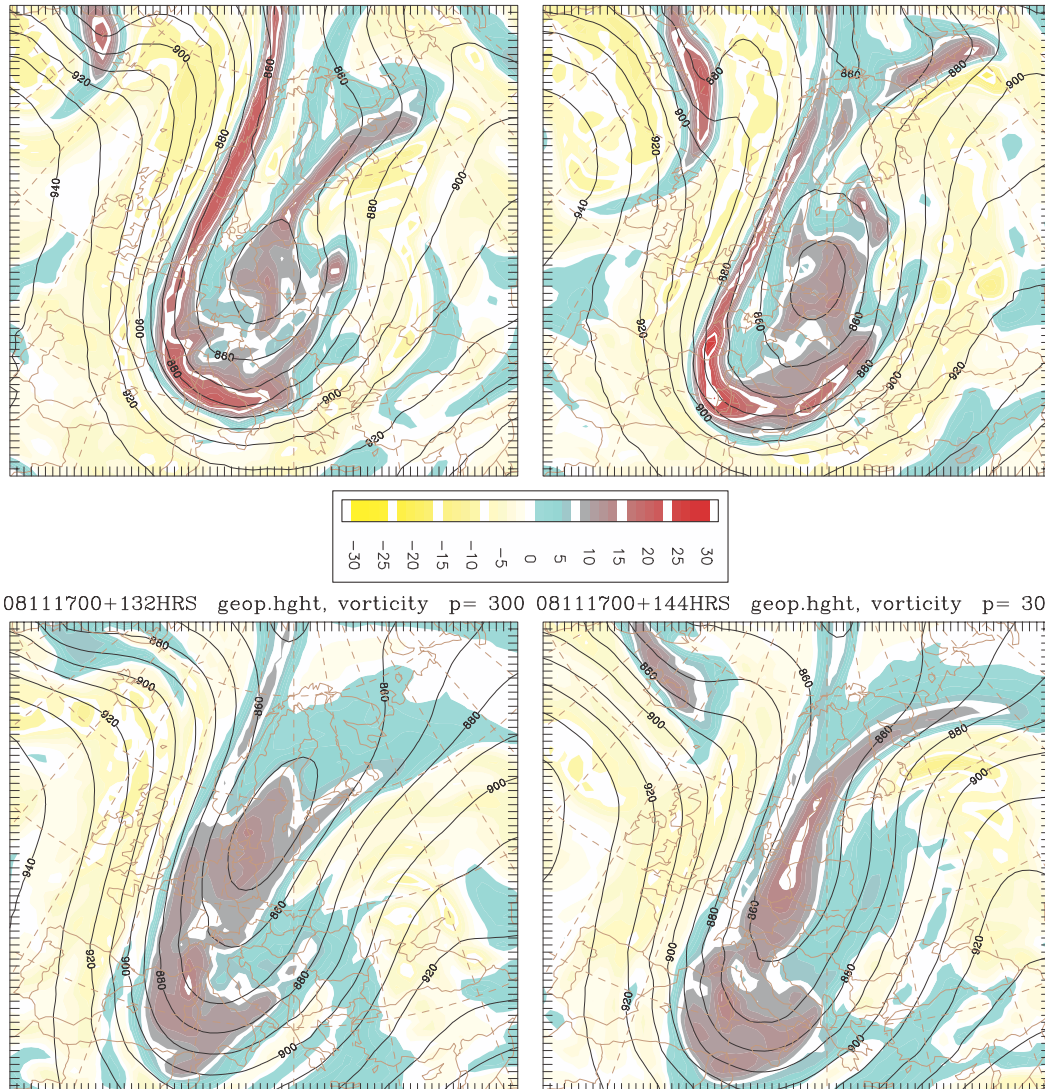

FIG. 8. As in Fig. 7, but for G7 resolution (~60 km).

conceptually consistent with the fact that the doming of isentropes in the vortex center reduces velocity gradients along coordinate surfaces on the cyclonic side of the jet, thereby lessening numerical dissipation of cyclonic vorticity below jet stream level.

7. Discussion

Evidence we have accumulated so far, although still largely anecdotal, suggests that FIM, using its native σ–θ coordinate, maintains the integrity of tropospheric vorticity filaments better with decreasing resolution than the same model using a σ–p coordinate adopted from NCEP's GFS. The evidence is based on three extensively analyzed cases and on a few more cases that were analyzed in less detail, all chosen from the 2008/09 boreal autumn and winter season.

Our reasoning for why different trends in the simulation of Rossby wave breaking in the two model versions are to be expected is straightforward. As argued earlier, *numerically* induced lateral mixing of fine structures generated by filament stretching is least detrimental to forecast accuracy in models whose mixing surfaces coincide with isosurfaces of potential buoyancy. FIM with its σ–θ grid represents such a model, at least above the lowest few kilometers of the atmosphere.

One seeming flaw of the above case study is the juxtaposition of two different vorticity fields—one, η_θ, based on differentiation at constant θ and one, η_p, based on differentiation at constant p. For this reason, comparison
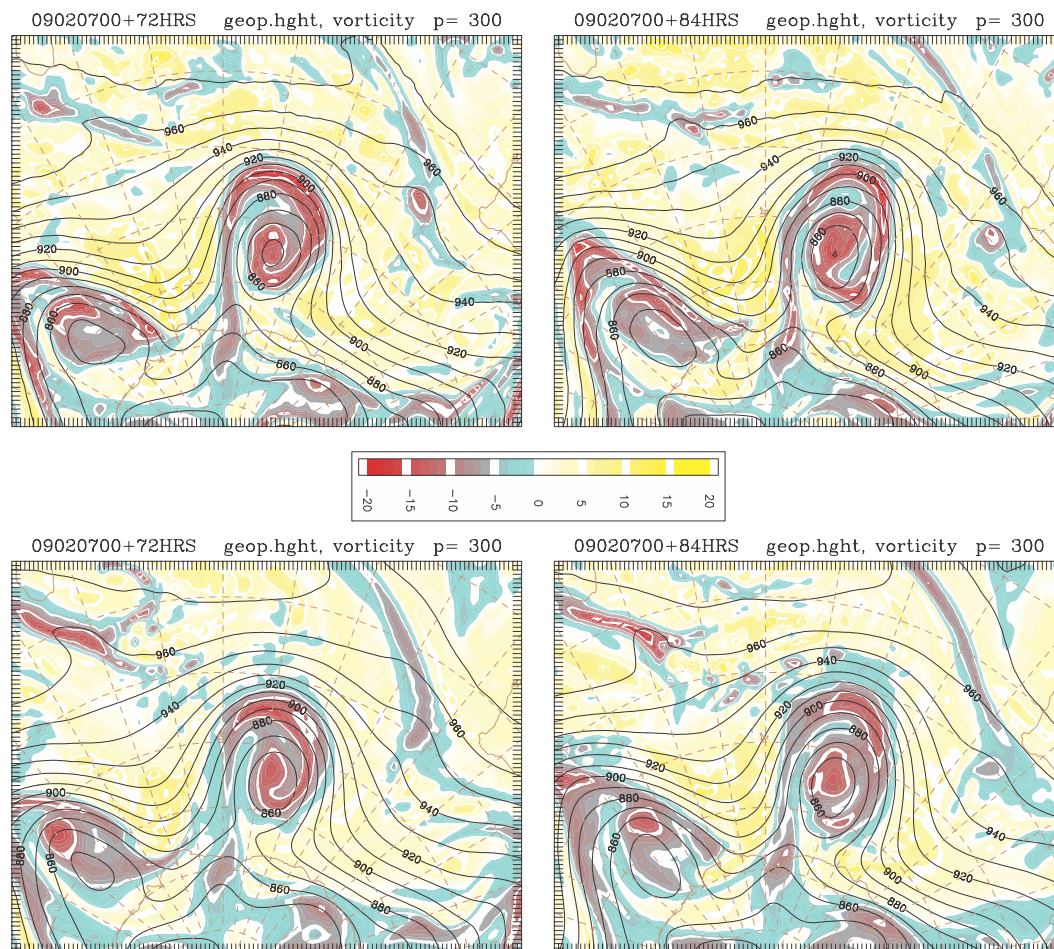
FIG. 9. 72- and 84-h forecasts of geopotential height (m) and vorticity (10^{-5} s^{-1}) at 300 hPa. The initial time is 0000 UTC 7 Feb 2009. (top) Forecasts based on native σ–θ grid. (bottom) FIM forecasts using GFS σ–p grid. The horizontal resolution is G8 ($\sim$30 km). Maps are cropped to show pieces of Madagascar and Australia for orientation.

of η_p with η_θ at the same spatial resolution is less meaningful than a comparison of the resolution dependence of either the η_θ or the η_p field by itself. We should therefore refrain from making conclusions such as "filament simulation in the σ–θ model at Gn resolution (where n is a natural number) compares accuracy-wise to simulation in the σ–p model at resolution G(n + 1)." On the other hand, it seems permissible to conclude that the deterioration of forecast accuracy with decreasing resolution is more rapid in the σ–p simulations than in the σ–θ simulations.

Even if quantitative information cannot be drawn from it, the exercise of plotting η_θ on p surfaces and comparing the resulting distribution with η_p plotted on the same surfaces is still informative. It illustrates the level of detail in a dynamically relevant tracer field that in an isentropic coordinate model, by virtue of coordinate surface orientation, is largely shielded from

nonphysical dissipation. The η_p field in the FIM version that features a σ–p coordinate, on the other hand, directly feels the impact of numerical dissipation and deteriorates with time accordingly. It remains to be seen, of course, whether some degree of deterioration might not be realistic. In other words, the question will have to be addressed as to whether the FIM equations should retain their present frictionless form or be amended by explicit mixing terms.

8. Conclusions

Material has been presented supporting the notion that flow-following or quasi-Lagrangian vertical coordinates are a viable alternative to the Eulerian coordinates commonly used in atmospheric circulation modeling. This point, of course, has been made before by members of the University of Wisconsin and University of California,
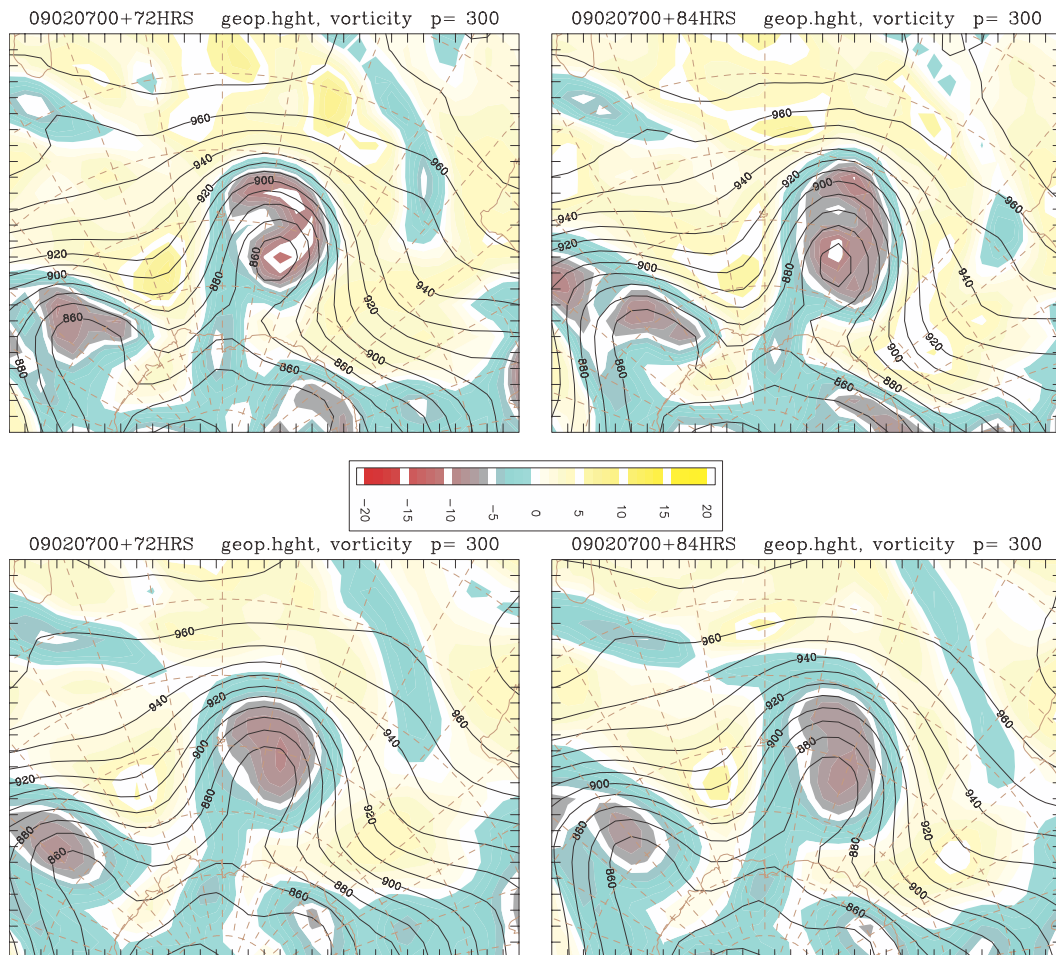
FIG. 10. As in Fig. 9, but for G6 resolution (~120 km).

Los Angeles, schools (Johnson 2000; Arakawa 2000; Randall et al. 2000, among others). The distinguishing feature of the present effort is that a quasi-isentropic coordinate model is being used on a routine basis for real-time, medium-range global weather prediction. In comparing FIM with the other two hybrid-isentropic models routinely used for weather prediction today, FIM differs from RUC (Benjamin et al. 2004) primarily in the use of an icosahedral global grid and from the University of Wisconsin global model (Schaak et al. 2004) in the use of an ALE coordinate.

Not all components of FIM are new. Physics parameterization codes were made available by the group responsible for the GFS at NCEP; FIM development efforts have greatly benefited from this collaboration. Likewise, the global fields used to initialize FIM are imported directly from NCEP.

This having been said, FIM is unique in having combined two novel approaches to numerical weather prediction: 1) icosahedron- and finite-volume-based horizontal discretization and 2) an entropy-based vertical coordinate. The present article documents the latter aspect in detail. In addition, material is presented suggesting that numerical diffusion attributed to the dispersive effects of the horizontal transport operators can be rendered less destructive by solving the dynamic equations in an isentropic coordinate system. The question of whether this by itself leads to improved forecast accuracy has not been explored in this article but will be the subject of future work.
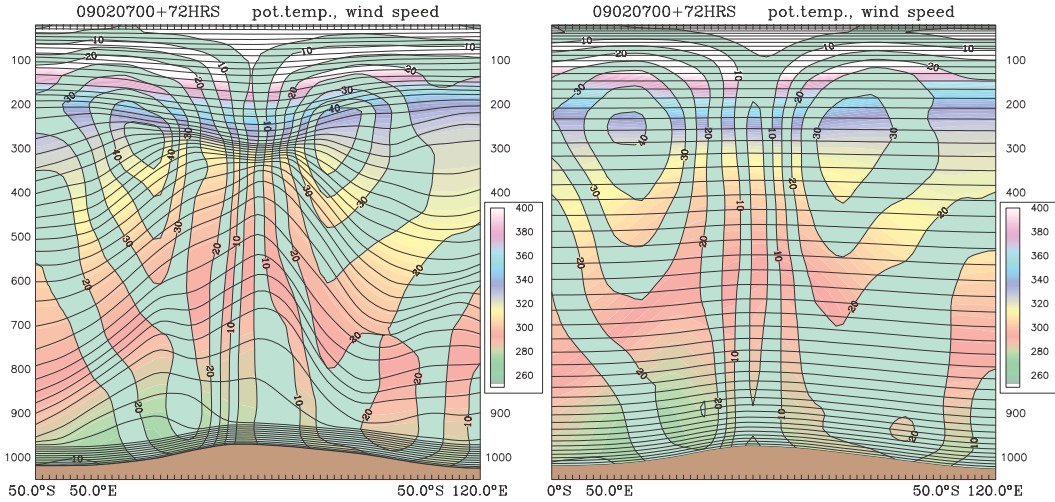
FIG. 11. Zonal cross sections at 50°S, 50°–120°E, extracted from 72-h G6 forecasts based on (left) σ–θ coordinate with 5-hPa minimum layer thickness and (right) GFS σ–p coordinate. Features shown in the sections are explained in Fig. 2.

APPENDIX A

Turbulent Vertical Mixing

The following is a simplified version of a numerical scheme developed by McDougall and Dewar (1998) for carrying out vertical mixing in fluid models whose vertical coordinate is a function of the diffused variable(s). The authors deal with the specific problem of mixing temperature and salinity in ocean models whose vertical coordinate is potential density (a function of both temperature and salinity), constrained to remain constant in each coordinate layer during mixing.

Here we address the simpler problem of solving the diffusion equation in an atmospheric column where there is only one diffused variable (potential temperature θ) doing double duty as vertical coordinate. The only variable capable of capturing the effects of thermal diffusion in this case is the thickness of coordinate layers.

The equations expressing conservation of mass and heat in a column, basically one-dimensional versions of the equations listed in the beginning, are

$$\frac{\partial}{\partial t}\left(\frac{\partial z}{\partial s}\right)_s + \frac{\partial}{\partial s}\left(\dot{s}\frac{\partial z}{\partial s}\right) = 0 \quad \text{and} \qquad (\text{A1})$$

$$\left(\frac{\partial \theta}{\partial t}\right)_s + \left(\dot{s}\frac{\partial z}{\partial s}\right)\frac{\partial \theta}{\partial z} = -\frac{\partial F_\theta}{\partial z}. \qquad (\text{A2})$$

The turbulent heat flux $F_\theta = \overline{w'\theta'}$ is usually parameterized as $F_\theta = -K\partial\theta/\partial z$, where θ is the resolved-scale potential temperature and K is a thermal diffusivity coefficient.

The flux form of (A2), obtained by combining (A1) and (A2), is

$$\frac{\partial}{\partial t}\left(\theta\frac{\partial z}{\partial s}\right) + \frac{\partial}{\partial s}\left(\dot{s}\frac{\partial z}{\partial s}\theta\right) = -\frac{\partial F_\theta}{\partial s}. \qquad (\text{A3})$$

The task at hand is to discretize the above equations for use in a model framework where stratification is represented by a piecewise-constant, stair-step θ profile. The discretization will be done by formally integrating the equations over individual stair steps.

If θ is to remain constant in each layer during the mixing process, F_θ must be vertically constant in each layer. If this were not the case, integrating (A2) over an individual layer would yield a nonzero right-hand side. Of the two terms on the left, the second one integrates to zero since $\partial\theta/\partial z = 0$ inside the layer. [The vertical mass flux ($\dot{s}\partial z/\partial s$) remains finite.] Hence, a nonzero rhs implies a nonzero tendency term $\partial\theta/\partial t$, which clashes with the stated requirement.

We conclude that for diffusion to leave a mark on the profile under the constraint $F_\theta = $ const in individual layers, F_θ must be allowed to vary from layer to layer. The implied infinite heat flux divergence at layer interfaces is consistent with the notion that air crossing an interface undergoes an instantaneous change in θ.

A simple centered finite-difference expression for the heat flux in layer n is

$$F_\theta^n = \frac{K^n}{2}\frac{\theta^{n+1} - \theta^{n-1}}{z^{n+1/2} - z^{n-1/2}}, \qquad (\text{A4})$$

where fractional superscripts indicate quantities defined on interfaces.

The central task is to determine the mass flux across layer interfaces, $(\dot{s}\partial z/\partial s)$. For this we integrate (A3) over an s interval representing an infinitesimal slab bracketing a layer interface. Since the tendency term drops out as ∂z approaches zero and the mass flux $(\dot{s}\partial z/\partial s)$ is continuous in the vertical, we obtain in the limit of zero slab thickness:

$$\left(\dot{s}\frac{\partial z}{\partial s}\right)^{n+1/2} = \frac{F_\theta^{n+1} - F_\theta^n}{\theta^{n+1} - \theta^n}. \quad (A5)$$

Expressions (A4) and (A5) encompass the sought-after solution to the problem of diffusing heat in a stair-step θ profile while maintaining θ in individual layers. Note that, in the absence of externally imposed heat fluxes, the column integral $\int \theta \, dz$ is conserved regardless of the physical and numerical approximations made in evaluating the heat flux in (A4).

The heat flux as approximated by (A4) becomes infinite in massless layers. To avoid division by zero, the denominator in (A4) must therefore be bounded away from zero. The parameter representing minimum layer thickness, together with K and the time step used in solving (A3), can be tuned to concentrate the effect of vertical diffusion almost entirely on very thin layers. We use the scheme in this mode as an alternative to the grid generator to avoid generating zero-thickness layers in the isentropic subdomain that may result from strongly layer-dependent diabatic forcing. The advantage of the present scheme over the grid generator is that it does not produce local deviations from target θ. Suitable parameter values are $1 \ m^2$ for the product of time step and mixing coefficient, and 2×10^{-3} m for the minimum thickness.

Ideally, vertical mixing should conserve the total heat content of the column, $\int c_v T \, dp$. From (11) we note that, in order to conserve total heat, the variable z in (A1)–(A5) must be replaced by a variable proportional to p^{1+k}. No other changes are required in the solution procedure, except that the mixing coefficient K in (A4) must be rendered dimensionally compatible with the new vertical coordinate.

In a similar vein, preservation of the total height $\int \theta \, d\Pi$ of the column during mixing can be achieved by using a variable proportional to p^k in place of z. Note that height preservation is incompatible with heat conservation.

APPENDIX B

Details of Minimum Thickness Enforcement

It is advisable to smooth out large lateral variations in layer thickness that typically occur where a hybridized coordinate layer transitions from the fixed-depth subdomain to the isentropic subdomain. These variations are created when, for a given k, the second argument in the *minimum* function of (10) is chosen in one grid column, while the first argument is chosen in a neighboring column. One way to smooth out the transition, short of exchanging information among neighboring grid columns, is to increase layer thickness in situations where the two arguments are of similar magnitude. This is the purpose of the "cushion" function originally introduced into hybrid-coordinate ocean modeling by Bleck and Boudra (1981) and later adapted for atmospheric use by Bleck and Benjamin (1993). Use of the cushion function entails replacing (10) by

$$\Pi_{k+1/2} = \min[\hat{\Pi}_{k+1/2}, \Pi_{k-1/2} - \text{cushn}(\hat{\Pi}_{k-1/2} - \Pi_{k+1/2}, \Delta\Pi_0)]. \quad (B1)$$

In the two extreme cases in which $\hat{\Pi}_{k-1/2} - \Pi_{k+1/2}$ is either large negative or large positive in comparison with $\Delta\Pi_0$, the cushion function is designed to replicate the functionality of (10). In other words, cushn(a, b) returns a if $a \gg b$, and it returns b if $-a \gg b$ $(b > 0)$. Between the two extremes, cushn varies smoothly, returning values as high as $2 \max(a, b)$. In many cases, this inflates a layer if its potential temperature is close to target, thereby softening the lateral interface height contrast between locations where the underlying layer is isentropic and where it is not.

If more effective interface smoothing in the σ–θ transition region is deemed necessary, a sideways-looking smoothing algorithm may be required.

At the time of this writing, the minimum thickness value $\Delta\Pi_0$ is set as follows:

1) A default value $\Delta\Pi_0(k)$ is specified for each layer k. Typical values (stated here in pressure units for easier reference) are 3 hPa in the bottom layer, gradually increasing to 10 or 20 hPa in layers above.

2) In an attempt to mimic the vertical spacing of conventional σ-coordinate layers, all $\Delta\Pi_0(k)$ are multiplied by the factor $(p_{srf} - p_{top})/(1000 \ \text{hPa} - p_{top})$, where p_{srf} is the surface pressure and p_{top} is the pressure level (400 hPa or smaller) where coordinate surfaces in a conventional σ-coordinate model cease to be terrain-following.

3) Starting in the lowest layer and moving up the column, $\Pi_{k+1/2}$ is compared with the lesser of (B1) and $\Pi_{1/2} - \sum_{n=1}^{k} \Delta\Pi_0(n)$. If it exceeds the minimum of these two values, it is replaced by that minimum. This is done recursively (i.e., altered interface values affect the inflation test in layers above).

4) The lowest layer *not* in need of inflation is labeled $k_{\sigma\theta}$; it marks the transition from the σ to the θ subdomain.

5) The upper interface of layer $k_{\sigma\theta}$ stays fixed by definition, but very thin isentropic layers qualifying for inflation based on (B1) can occur higher up in the atmosphere. To keep these from unnecessarily being inflated, the value $\Delta\Pi_0(k)$ is reduced in layers $k_{\sigma\theta} + 1, \ldots, k_{\sigma\theta} + 4$ by the factors 0.4, 0.2, 0.1, and 0.05, respectively. The factor 0.05 is also used in layers $k > k_{\sigma\theta} + 4$.

Discretization of a continuous profile $\theta(\Pi)$ in terms of a stair-step profile is not unique, because stair steps can be broken into smaller steps or combined into bigger ones without violating any continuity or conservation principle. This ambiguity can lead to computational modes in the vertical layer structure, leading to the gradual disappearance of, say, odd-numbered layers accompanied by a thickening of even-numbered ones. Initial experiments with FIM indeed revealed a propensity for amplifying this mode. To suppress it, a special algorithm has been added to the grid generator.

The algorithm scans each grid column for sequences of 5 $\Delta\Pi$ values, numbered $\Delta\Pi_1, \ldots, \Delta\Pi_5$, that satisfy the following three conditions:

$$\Delta\Pi_1 < \Delta\Pi_2,$$
$$\Delta\Pi_5 < \Delta\Pi_4, \quad \text{and}$$
$$\Delta\Pi_3 < \min(\Delta\Pi_2, \Delta\Pi_4).$$

If all three conditions are met, layer 3 is inflated by drawing mass from both layers 2 and 4 such that 1) the column integral of θ is conserved and 2) $\Delta\Pi_3^{\text{new}} = \min(\Delta\Pi_2^{\text{new}}, \Delta\Pi_4^{\text{new}})$. Requirement 1 leads to the constraint

$$\frac{\Delta\Pi_2 - \Delta\Pi_2^{\text{new}}}{\Delta\Pi_4 - \Delta\Pi_4^{\text{new}}} = \frac{\theta_4 - \theta_3}{\theta_3 - \theta_2},$$

which may put a limit on the mass transfer stipulated by requirement 2. The resulting interface displacements are added to those associated with the primary regridding process.

Suppression of the layer-thickness computational mode improves the performance of the column physics parameterization scheme, which has been found to be sensitive to large variations in layer thickness.

REFERENCES

Arakawa, A., 2000: Future development of general circulation models. *General Circulation Model Development,* D. A. Randall, Ed., International Geophysical Series, Vol. 70, Academic Press, 721–780.

——, and V. R. Lamb, 1981: A potential enstrophy- and energy-conserving scheme for the shallow water equations. *Mon. Wea. Rev.,* **109,** 18–36.

——, and S. Moorthi, 1988: Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.,* **45,** 1688–1707.

Benjamin, S., G. Grell, J. Brown, T. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.,* **132,** 473–494.

Bleck, R., 1974: Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Wea. Rev.,* **102,** 813–829.

——, 1978a: Finite difference equations in generalized vertical coordinates. Part 1: Total energy conservation. *Beitr. Phys. Atmos.,* **51,** 360–372.

——, 1978b: On the use of hybrid vertical coordinates in numerical weather prediction models. *Mon. Wea. Rev.,* **106,** 1233–1244.

——, 1984: An isentropic coordinate model suitable for lee cyclogenesis simulation. *Riv. Meteor. Aeronaut.,* **44,** 189–194.

——, 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modell.,* **4,** 55–88.

——, and D. B. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.,* **11,** 755–770.

——, and ——, 1986: Wind-driven spin-up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Geophys. Res.,* **91C,** 7611–7621.

——, and S. Benjamin, 1993: Regional weather prediction with a model combining terrain-following and isentropic coordinates. Part I: Model description. *Mon. Wea. Rev.,* **121,** 1770–1785.

Boris, J. P., and D. L. Book, 1973: Flux-corrected transport. I. Shasta, a fluid transport algorithm that works. *J. Comput. Phys.,* **11,** 38–69.

Colella, P., and P. Woodward, 1984: The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Comput. Phys.,* **54,** 174–201.

Deaven, D. G., 1976: A solution for boundary problems in isentropic coordinate models. *J. Atmos. Sci.,* **33,** 1702–1713.

Dritschel, D. G., and L. M. Polvani, 1992: The roll-up of vorticity strips on the surface of a sphere. *J. Fluid Mech.,* **234,** 47–69.

Eliassen, A., and E. Raustein, 1968: A numerical integration experiment with a six-level atmospheric model with isentropic information surface. *Meteor. Ann.,* **5,** 429–449.

Friend, A. L., D. Djuric, and K. C. Brundidge, 1977: A combination of isentropic and sigma coordinates in numerical weather prediction. *Beitr. Phys. Atmos.,* **50,** 290–295.

Hirt, C. W., A. A. Amsden, and J. L. Cook, 1974: An arbitrary Lagrangian-Eulerian computing method for all flow speeds. *J. Comput. Phys.,* **14,** 227–253.

Hsu, Y.-J. G., and A. Arakawa, 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Mon. Wea. Rev.,* **118,** 1933–1959.

Johnson, D. R., 2000: Entropy, the Lorenz energy cycle, and climate. *General Circulation Model Development,* D. A. Randall, Ed., International Geophysical Series, Vol. 70, Academic Press, 659–720.

——, and Z. Yuan, 1998: The development and initial tests of an atmospheric model based on a vertical coordinate with a smooth transition from terrain following to isentropic coordinates. *Adv. Atmos. Sci.,* **15,** 283–299.

——, T. H. Zapotocny, F. M. Reames, B. J. Wolf, and R. B. Pierce, 1993: A comparison of simulated precipitation by hybrid isentropic-sigma and sigma models. *Mon. Wea. Rev.,* **121,** 2088–2114.

Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.,* **102,** 509–522.

Kleinschmidt, E., 1950: Über Aufbau und Entstehung von Zyklonen (1. Teil). *Meteor. Rundsch.,* **3,** 1–6.

Konor, C. S., and A. Arakawa, 1997: Design of an atmospheric model based on a generalized vertical coordinate. *Mon. Wea. Rev.,* **125,** 1649–1673.

Lee, J.-L., and A. E. MacDonald, 2009: A finite-volume icosahedral shallow-water model on a local coordinate. *Mon. Wea. Rev.,* **137,** 1422–1437.

Lin, S. J., 2004: A vertically Lagrangian finite-volume dynamical core for global models. *Mon. Wea. Rev.,* **132,** 2293–2307.

McDougall, T. J., and W. K. Dewar, 1998: Vertical mixing and cabbeling in layered models. *J. Phys. Oceanogr.,* **28,** 1458–1480.

McIntyre, M. E., and T. N. Palmer, 1985: A note on the general concept for wave braking of gravity and Rossby waves. *Pure Appl. Geophys.,* **123,** 964–975.

Montgomery, R. B., and A. F. Spilhaus, 1941: Examples and outline of certain modifications in upper-air analysis. *J. Aeronaut. Sci.,* **8,** 276–283.

Phillips, N. A., 1957: A coordinate system having some special advantages for numerical forecasting. *J. Meteor.,* **14,** 184–185.

Pierce, R., D. Johnson, F. Reames, T. Zapotocny, and B. Wolf, 1991: Numerical investigations with a hybrid isentropic-sigma model. Part I: Normal mode characteristics. *J. Atmos. Sci.,* **48,** 2005–2024.

Randall, D. A., T. D. Ringler, and R. Heikes, 2000: Global atmospheric modeling using a geodesic grid and an isentropic vertical coordinate. *General Circulation Model Development,* D. A. Randall, Ed., International Geophysical Series, Vol. 70, Academic Press, 509–536.

Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *J. Phys. Oceanogr.,* **12,** 1154–1158.

Rossby, C.-G., 1940: Planetary flow patterns in the atmosphere. *Quart. J. Roy. Meteor. Soc.,* **66** (Suppl.), 68–87.

——, and Collaborators, 1937: Isentropic analysis. *Bull. Amer. Meteor. Soc.,* **18,** 201–209.

Sadourny, R., 1975: The dynamics of finite-difference models of the shallow-water equations. *J. Atmos. Sci.,* **32,** 680–689.

Schaak, T. K., T. H. Zapotocny, A. J. Lenzen, and D. R. Johnson, 2004: Global climate simulation with the University of Wisconsin global hybrid isentropic model. *J. Climate,* **17,** 2998–3016.

Starr, V. P., 1945: A quasi-Lagrangian system of hydrodynamical equations. *J. Meteor.,* **2,** 227–237.

Takacs, L. L., 1985: A two-step scheme for the advection equation with minimized dissipation and dispersion errors. *Mon. Wea. Rev.,* **113,** 1050–1065.

Uccellini, L. W., D. R. Johnson, and R. E. Schlesinger, 1979: An isentropic and sigma coordinate hybrid numerical model. Part I: Model development and some initial tests. *J. Atmos. Sci.,* **36,** 390–414.

van Leer, B., 1974: Towards the ultimate conservative difference scheme. II: Monotonicity and conservation combined in a second order scheme. *J. Comput. Phys.,* **14,** 361–370.

Webster, S., J. Thuburn, B. J. Hoskins, and M. J. Rodwell, 1999: Further development of a hybrid-isentropic GCM. *Quart. J. Roy. Meteor. Soc.,* **125,** 2305–2331.

Zapotocny, T., D. Johnson, F. Reames, R. Pierce, and B. Wolf, 1991: Numerical investigations with a hybrid isentropic-sigma model. Part II: The inclusion of moist processes. *J. Atmos. Sci.,* **48,** 2025–2043.

——, ——, and ——, 1994: Development and initial test of the University of Wisconsin global isentropic-sigma model. *Mon. Wea. Rev.,* **122,** 2160–2178.

Zhu, Z., J. Thuburn, B. J. Hoskins, and P. H. Haynes, 1992: A vertical finite-difference scheme based on a hybrid σ–θ–p coordinate. *Mon. Wea. Rev.,* **120,** 851–862.